Sensitivity and Uncertainty Analysis

in Hydrologic Simulation Modeling

of the

South Florida Water Management District

Report of the Workshop on

Reduction of Uncertainties in Regional Hydrologic Simulation Models

(SFWMM and NSM)

January 18 - 19, 1994

Prepared by:

Daniel P. Loucks and Jery R. Stedinger

School of Civil and Environmental Engineering

Hollister Hall, Cornell University, Ithaca, NY 14853

March 1, 1994

Table of Contents

	Page
Introduction	2
Issues, Concerns, and Terminology	3
Hydrologic Modeling and Output Uncertainty	4
Decisions in the Face of Uncertainty	9
Causes of Variability and Uncertainty in Model Output	10
Impacts of Input Uncertainty on Model Output	12
Sensitivity Analysis	16
Modeling and Input Data Errors	16
Users of Sensitivity Analyses	18
Sensitivity Analysis Philosophies	19
Deterministic Sensitivity Analysis	21
First-Order Uncertainty Analysis	27
Approximate First-Order Uncertainty Analysis	28
Expected value of additional information	31
Fractional Factorial Design Method	32
Monte Carlo Sampling Methods	34
Simple Monte Carlo Sampling	37
Standardized Monte Carlo analysis	41
Latin Hypercube Sampling	42
Specialized Methods for Water Quality Models	45
Calibration of Hydrologic Simulation Models	46
Advantages of Automatic Calibration Packages	46
Statistical Descriptions of Parameter Precision	47
Sensitivity Analysis Program	48
Displaying Model Performance Indices	52
Measures to Reduce Risk and Uncertainty	54
Distinguishing Differences between Performance Indicator Distributions	55
Conclusions	57
References	59

Introduction

This report summarizes the results of a workshop on risks and uncertainties in hydrologic modeling held by the South Florida Water Management District. The two-day workshop, on 18 and 19 January, 1994, included representatives from various Federal Agencies (including the US Army Corps of Engineers, the US Environmental Protection Agency, the US Geological Survey, and the National Park Service) as well as individuals from the District. The workshop focused on two hydrologic models currently in use by the District, the South Florida Water Management Model (SFWMM) and the Natural Systems Model (NSM). This report is based on the discussions held during the workshop. It is intended to serve as a guide for identifying, quantifying and reducing the risks and uncertainties associated with various system performance indicator values that are derived from those hydrologic simulation models.

Methods for quantification of sensitivity and uncertainties in the output of simulation models have been developed and refined over the past two decades (Beck, 1987; DeGarmo et al., 1993). Many of these methods are not very useful or practicable in large-scale hydrologic simulations, such as performed when using either SFWMM or NSM. The workshop provided an opportunity for participants to express their opinions concerning the relative importance of various sources of risk and uncertainty, and their preferences for methods that can quantify or reduce that risk and uncertainty. This report summarizes the workshop discussions, and outlines in greater detail some of the methods that seem most appropriate.

Workshop Goals. The workshop had a number of goals. These goals reflect the interests and responsibilities of different individuals, departments and agencies. The goals included:

- Achievement of a common understanding of the issues, concerns, and terminology concerning risk and uncertainty in model outputs.
- Identification of the causes of variability, risk, and uncertainty in model outputs.
- Ranking of model parameters with respect to their impact on the uncertainty associated with the values of model outputs and system performance indicators.
- Identification of ways to determine and display the risk or uncertainty associated with model output or system performance indicators.

- Identification of measures to reduce the uncertainty associated with values of system performance indicators, including field investigations.
- Identification of methods for distinguishing between significantly different probability distributions of system performance indicator values.
- Identification of efficient ways to calibrate hydrologic simulation models that will produce an accurate and robust model (i.e. a model whose parameter values can be varied over wide ranges without requiring model recalibration).

The remainder of this report addresses each of these goals. It begins with a general discussion of risk and uncertainty, and the causes of risk and uncertainty in model output predictions. It then examines some ways of measuring or quantifying such risk and uncertainty, concentrating on methods that seem most relevant, or practical, for the District. These methods involve not only calculations of ranges of possible, or likely, output and system performance indicator variable values, but also displaying those ranges in ways meaningful to model users, managers and decision makers. The report concludes with some discussion of model calibration issues, statistical interpretation techniques, and measures that could be taken to reduce risk and uncertainty.

Issues, Concerns, and Terminology

Outcomes or events that cannot be predicted with certainty are often called risky or uncertain. Some individuals draw a special and interesting distinction between **risk** and **uncertainty**. In particular, the term risk is often reserved to describe situations for which probabilities are available to describe the likelihood of various events or outcomes. If probabilities of various events or outcomes cannot be quantified, or responsible individuals are unwilling to provide estimates of the probabilities, some would say the problem is then one of uncertainty, and not of risk. When probabilities are measurable, risk is called **objective risk.** If the probabilities are based solely on human judgement, the risk is called **subjective risk.** It was clear from the discussions in the workshop that such distinctions between objective and subjective risk, and between risk and uncertainty, serve no useful purpose to those developing and using the District's models. Likewise the distinctions are often unimportant to those who should be aware of the risks or uncertainties associated with system performance indicator values. Hence in this report the words risk and uncertainty are used interchangeably.

Inadequate information is inherent in any future-oriented planning effort. Risk and uncertainty stem from inadequate information and incorrect assumptions, as well as from the variability of natural processes. One objective of water managers in the District is to identify the sensitivity of system performance indicator values to various sources of errors that cause risk or uncertainty, to reduce this level of risk and uncertainty to the extent practicable, and to communicate the residual uncertainties clearly so that decisions can be made with this knowledge and understanding. Alternative ways of meeting these objectives were among the subjects addressed at the workshop and are described in the program developed in this report.

Technical experts who analyze risk, managers and decision makers who must manage risk, and the public who must live with risk and uncertainty, have different information needs and attitudes regarding risk and uncertainty. It is clear that information needs differ among those who model or use models, those who make substantial investment decisions, and those who are likely to be impacted by those decisions. Hence attention needs to given to the communication of risk and uncertainty outside of the modeling group. This workshop report addresses primarily model developers and model users.

Hydrologic Modeling and Output Uncertainty

The SFWMM and NSM hydrologic models are used by engineers in the District to predict possible hydrologic impacts of alternative water management policies under a variety of hydrologic inputs. These models are static in the sense that structural and demand conditions are fixed for a certain time period, and then a time series of hydrologic events are simulated. This yields a time series of model outputs (flows, volumes, losses, etc.) at approximately 2,000 locations throughout the District. These model outputs are then converted to a number of system performance indicators that reflect system reliability as well as the durations and extents of undesirable or "failure" conditions at various sites in the region.

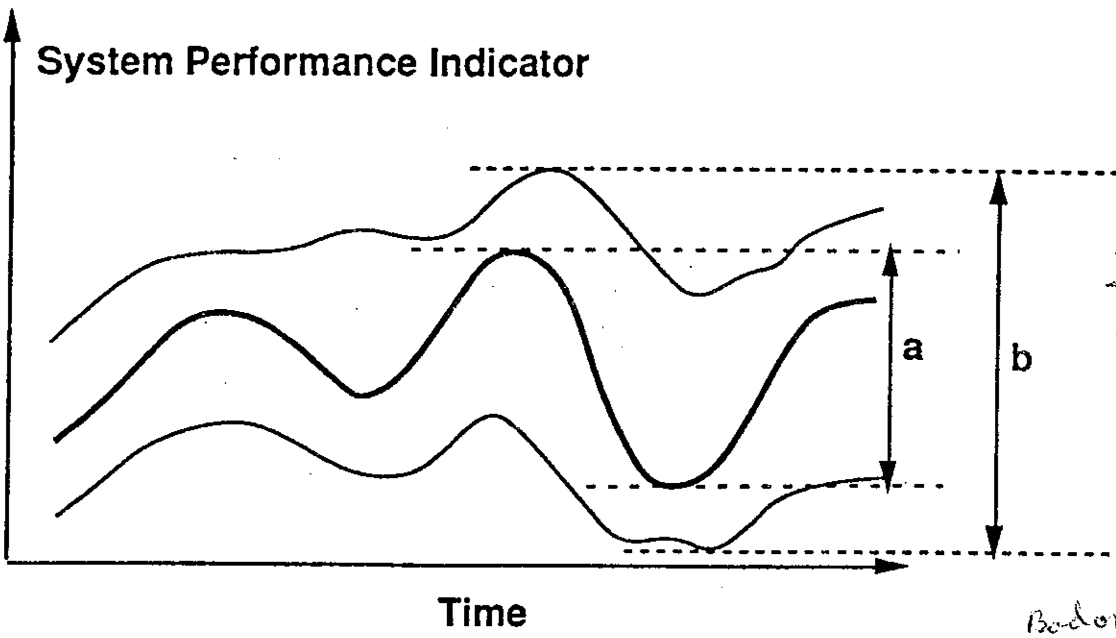
The use of predictions obtained with these hydrologic simulation models requires a quantitative understanding of the sensitivity of model predictions to various sources of uncertainties. Model users should be interested in the range of possible values of each model output as well as their most likely values. Both effect the values of the system performance indicators and their reliabilities. Decisions regarding investments in new control structures or the operation of existing facilities may depend on the level of confidence associated with the predicted system performance indicators. If so, attention should be paid to the levels of risk and uncertainty associated with each system performance indicator.

Errors that affect precision of predictions. The values of each simulated variable at each specific location in the District will vary over time. Variability of model output is a direct result of variability of model input (hydrologic and meteorologic data). However, the extent of the variability, and the lower and upper limits of that variability, will also be affected by errors in:

- the hydrologic and meteorologic inputs;
- the values of parameters associated with various natural processes, such as evapotranspiration or overland and subsurface flow, or water quality constituent transport, growth or decay and transformation;
- the initial boundary conditions;
- the model structure and/or numerical solution techniques;
- the model operating policies involving water supply and flood control operations, canal stage maintenance, water control structure operation, well-field withdrawals and water shortage policies - all in comparison to what is actually done in the field; and possibly others as well.

There is also uncertainty with respect to human behavior and reaction related to particular outcomes and their likelihoods, i.e., to their risks and uncertainties. As important as risks and uncertainties associated with human reactions are to particular outcomes, they are not part of the hydrologic models themselves. Thus they will not be addressed in this report.

Natural variability and the effect of possible errors. Figure 1 illustrates the distinction between the variability of a system performance indicator due to natural hydrologic or meteorologic data variability, and the extended range of variability due to the risk or uncertainty associated with any combination of the errors listed above. This extended range is addressed in this report.



b may be smaller than a due to data checks that throw out largest values, Estimation of Bad or Missing data and Model ~~errors~~. Sometimes missing [illegible]

Figure 1. Time-series of model output or system performance showing variability over time. Range "a" results from the natural variability of hydrologic and meteorological input data over time. Assuming range "a" represents an error-free range, the extended range "b" results from the variability of natural input data as well as from errors in input data measurement, parameter value estimation, model structure and model solution algorithms. The extent of this range will depend on the confidence level associated with that range.

If the errors that caused the extended range of model output or system performance indicator values shown in Figure 1 were known, one could actually construct graphs similar to those shown in Figure 1. Since most of these errors are not known, what can occur in practice is a time-series of system performance indicator values that can range anywhere within or even outside the extended range, assuming the confidence level of that extended range is less than 100%. Figure 2 illustrates this point.

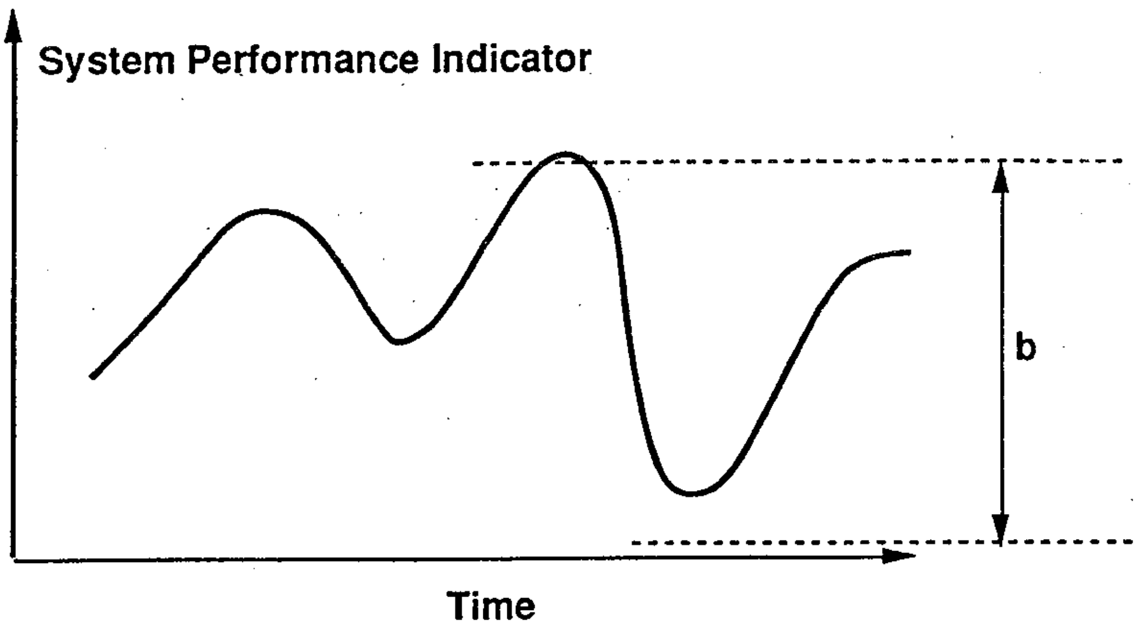


Figure 2. Typical time series of model output or system performance indicator values that are the result of input data variability and errors in input data measurement, parameter value estimation, model structure and model solution algorithms. Note that the time series may even contain values outside the range "b" defined in Figure 1 if its confidence level is less than 100%.

Calibration Precision and Prediction errors. It is worthwhile to explore the transformation of uncertainties in model inputs and parameters into uncertainty in model output. In this context, the District has data on system operations for some 50 years, and details records for the last 25 or so. These historical records along with data on system characteristics, including land cover, elevations and dimensions of canals and control structures, comprises the knowledge base from which the District can project system performance under different conditions. Those differences may include different sequences of precipitation and potential evaporation, changes in operating policies resulting in different stages and flows, or different land use and changes in physical facilities including canals, levies and gates.

Figure 3 illustrates this issue. If asked how the system would operate with meteorology sequences, operating policies, and demands very similar to those in the historical data base, the model should be able to interpolate within the available knowledge base to provide a fairly precise estimate. Still that estimate will not be perfect, because our ability to reproduce current and recent operations is not perfect, though it should be fairly good.

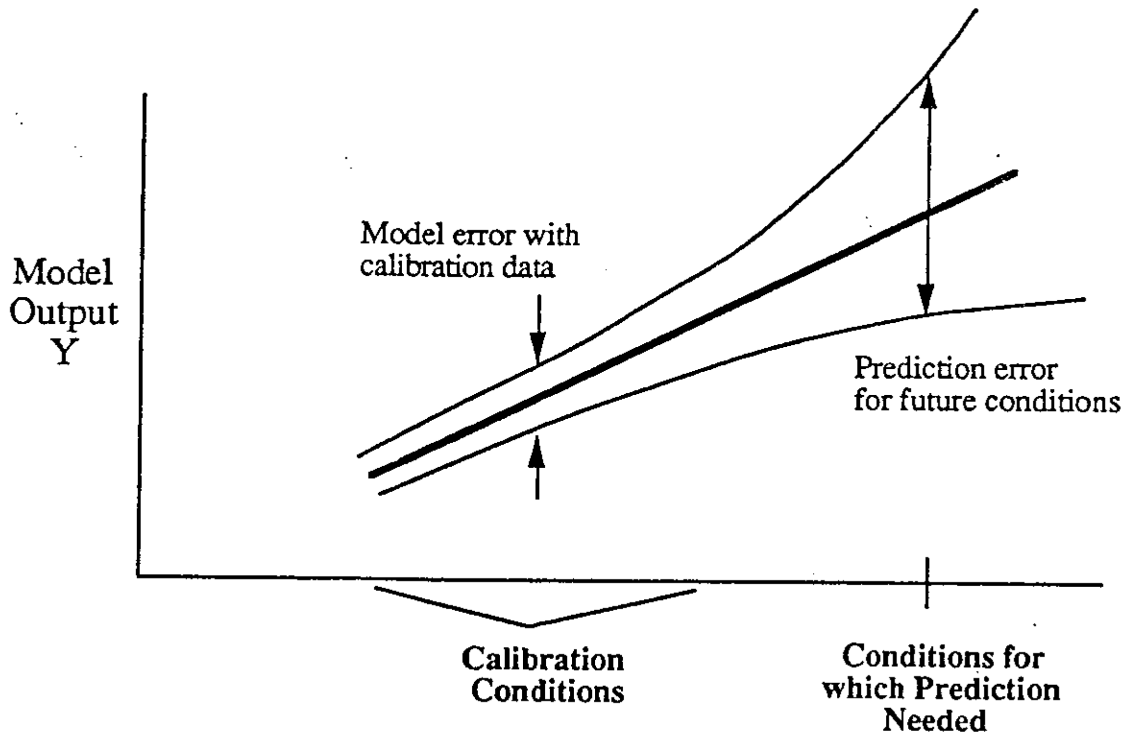


Figure 3. The precision of model predictions is affected the difference between the conditions of interest and the conditions for which the model was calibrated.

When asked to predict system performance for situations very different from those in the historical knowledge base, such predictions become much less precise for two reasons. First, our description of the characteristics of those different situations may be imprecise (i.e. what are appropriate elevations in the region and flow boundary conditions for the natural system model). Second, the knowledge base is insufficient for calibrating model parameters so that we can not reliably predict how the system will operate with stages and flow levels unlike those that have been experienced historically. The flow and quality models contain physically-motivated aggregated descriptions of hydraulic processes. The parameters of those models have been calibrated using the available historical data sets, which are also imperfect. The more conditions of interest are unlike those in the historical knowledge base, the less confidence one should have that the model is providing a reliable description of systems operation (Sorooshian et al., 1983; Gan and Burges, 1990). Clearly a sensitive analysis needs to consider how well the model can replicate current operations, and how similar the target conditions are to those described in the knowledge base. The greater the required extrapolation from what has been observed, the greater will be the importance of parameter and model uncertainties.

The relative and absolute importance of different parameters in different locations (evapotranspiration, hydraulic characteristics, seepage rates under levies) will depend on the system performance indicators of interest and on what regions they focus. Seepage rates may have a very large local effect, but a small global effect. Changes in system-wide evapotranspiration rates will clearly impact system-wide flows, including those flows entering the Everglades.

For these reasons the precision of model projections and the relative importance of errors in different parameters will depend upon:

(1) the precision with which the model can reproduce observed conditions,

(2) the difference between the conditions predicted and the historical experience included in the knowledge base, and

(3) the system performance characteristics of interest.

Possible errors in input data measurement, parameter value estimation, and model structure and model solution algorithms, are sources of risk and uncertainty. Their impacts on system performance indicator values, and what might be done to reduce them, were discussed in the workshop. The remainder of this report summarizes discussions aimed at identifying a reasonable program to quantify and reduce the range of uncertainty of various system performance indicator values resulting from possible errors. The end result, however, will not be a complete reduction in risk or uncertainty. Decisions will still have to be made in the face of uncertainty.

Decisions in the Face of Uncertainty

The District will certainly want to seek solutions to its water resources and environmental problems that are **flexible** and **resilient** in the face of uncertainty relating to hydrologic parameters, physical characteristics of the watersheds, and the needs and wants of future generations. **Incremental strategies** attempt to solve short-term problems, while gaining new knowledge and making progress toward long-term issues making irreversible commitments. Phased construction is an example of this kind of planning. Likewise, tight regulations and targets should not be adopted if it is very possible they cannot be met. When uncertainties associated with system operation under a new operating regime are large, one should seek flexible designs, and anticipate the need to make changes and improvements as experience and new information accumulate. When predictions are highly unreliable, responsible managers should consider alternative hypotheses and system performance for each reasonable scenario. They should favor actions that are robust (e.g., good under a wide range of situations), gain information, probe and experiment, monitor results to provide feedback for the next step, update assessments and modify policies in the light of new information, and avoid irreversible actions and commitments.

Causes of Variability and Uncertainty in Model Output

Natural meteorological variability and data errors. The main source of model output value variability is the natural variability in the hydrologic and meteorologic input series. Periods of normal precipitation and temperature have been interrupted by periods of extended drought and intense meteorological events such as hurricanes. There is no reason to think such events will not continue to occur. Sensitivity analysis can focus on errors in predictions made by deterministic hydrologic models for specified sets of time-series data describing precipitation, temperature and other exogenous forces across and on the border of the regions being studied.

Time series input data are usually historical. The time-series values typically describe historical conditions including droughts and wet periods. What is distinctive about natural uncertainty, as opposed to errors due to modeling limitations, is that natural variability in meteorological forces cannot be reduced by improving the model's structure, increasing the resolution of the simulation, or by better calibration of model parameters.

Errors result if meteorological values are not measured or recorded accurately, or if mistakes are made in the generation of computer data files. Furthermore, there is no assurance the statistical properties of historical data will accurately represent the statistical properties of future data. Actual future precipitation and temperature scenarios will be unlike those in the past, and this uncertainty in many cases may be much larger than the uncertainty due to input parameter error. In addition, the affects of uncertainties in the parameter values used in stochastic generation models are often much more significant than the affects of using different stochastic generation models (Stedinger and Taylor, 1982).

Parameter errors. A second main source of uncertainty in model output results from incorrect estimates of various parameter values used in the hydrologic models. These parameters include constants associated with functions predicting evapotranspiration or overland and subsurface flow, or water quality constituent transport, growth or decay and transformation. [See the section on calibration of models.]

Boundary conditions. A third source of uncertainty in model output results from errors in specifying the boundary conditions. These boundary conditions are fixed, or they can be variable. However, because they are not being computed based on the state of the system, they are an input that could be in error. These errors can affect the model output, especially in the vicinity of the boundary in each time step of the simulation.

Model structural and computational errors. A fourth source of uncertainty in model output can result from errors in the model structure compared to the real system, and errors resulting from numerical methods carried out in the simulation. Increasing model complexity to more closely represent the complexity of the real system may not only add to the cost of data collection, but also introduce even more parameters, and even more errors in model output. It is not an easy task to judge the appropriate level of model complexity, and to estimate the resulting levels of uncertainty associated with various assumptions regarding model structure and solution methods.

Uncertainty in Human Activities. The fifth main source of risk and uncertainty can result from unanticipated changes in human activities, demands, and impacts. An example of this is the deviation from standard or published operating policies by operators in the field, as compared to what is incorporated into the hydrologic models. Comparing field data with model data for model calibration may yield incorrect calibrations if operating policies actually implemented in the field differ significantly from those built into the simulation models.

Surprises. The District may also want to consider system vulnerability to undesirable environmental surprises. What havoc might an introduced species (like the zebra mussel invading the Great Lakes of the US) have in south Florida? Might some introduced disease suddenly threaten key plant or animal species? Might management plans have to be restructured to address the survival of some endangered species (such as salmon in the northwest), or some small community of plants and animals? Such uncertainties are hard to anticipate when by their nature they are truly surprises. But surprises should be expected and hence system flexibility should be maintained to deal with changing management demands, objectives, and constraints.

What humans will want to achieve in the future may not be the same as what they want today, and this is clearly a source of uncertainty. Human activities and demands for water resources and environmental amenities are driven by complex social and economic processes. Some of these processes reflect local concerns and activities, but population migration and many economic activities and social attitudes can reflect national and international trends. **Sustainability** is a concept in current vogue describing management approaches that attempt to use resources in current time frames without foreclosing options or diminishing the standard of living enjoyed by future generations.

The impact of variations in human activities at the boundaries and within the area managed by the District can again be described by scenarios for inclusion in sensitivity studies. It is important that careful attention go into the development of these alternative scenarios so that they realistically capture the forces that the system may face. The history of systems studies are full of examples where the issues studied were rapidly overwhelmed by much larger social forces resulting from relocation of major economic activities, an oil embargo, changes in national demand for natural resources, economic recession, or even war. One thing is sure, the future will be different than the present and the past.

Impacts of Input Uncertainty on Model Output

One source of considerable risk or uncertainty stems from the estimation of parameters in the hydrologic simulation model. Some parameters have physical meaning, others are simply coefficients that are determined during model calibration. During the workshop, many input data and processes algorithms were identified as significant to model prediction precision and are listed below. This section also describes the key model output and system performance indicators that were identified and discussed at the workshop.

Input Data. Model predictions are driven in large part by description of the input meteorology, parameters that allow translation of inputs into modelled hydrologic fluxes, and land cover and subsurface descriptions. Critical input data include:

- Hydrologic and meteorological time-series input data including daily rainfall, daily Penman-Monteith Reference ET, actual or predicted well pumpage, and daily estimated LEC developed area irrigation demands.
- Evapotranspiration parameters including those based on time series of solar radiation, wind, and humidity; land cover (crop) coefficients (K_{veg}); "open water' coefficients (K_p); and root zone depths (SRZ and DRZ).
- Overland flow parameters including Manning's n and surface detention depths for various land uses or land cover types.
- Infiltration parameters including infiltration rates and soil storage coefficients.
- Groundwater flow parameters including aquifer depths, hydraulic conductivities or transmissivities, depth of unsaturated zone and its moisture content, well withdrawals, and natural recharge rates.
- Canal-aquifer hydraulic conductivity coefficient, and levee seepage rates.
- Land elevations.
- Size of grid cell (model spatial resolution).
- Duration of simulation time step (model temporal resolution).
- Flow control structure operation policy.
- Historical data used for history matching, including water levels and structure discharges.
- Demand and boundary condition estimates.

Process Algorithms. Hydrologic process algorithms are also a source of error, and hence are also a source of imprecision in model predictions. Critical processes include:

- Evapotranspiration from land and water surfaces.
- Surface or Overland flow (runoff).
- Infiltration and percolation.
- Groundwater flow.
- Canal flow.
- Canal-aquifer interactions.

Model Outputs. Each of the above sources of risk and uncertainty can affect the values of some of the model output variable values. Hydrologic model outputs include:

- canal stages,
- structure discharges,
- water levels,
- ponding depths,
- evapotranspiration,
- overland flow,
- groundwater flow,
- canal water budgets, and
- subbasin water budgets.

Water use
and portion Met

Potential
vs
Actual

System Performance Indicators. Hydrologic model outputs are themselves a measure of system performance. These outputs can also be converted to additional, and often more meaningful, system performance indicator values. System performance indicators can apply to any grid-cell or to a group of grid cells (which can include the entire District). The system performance indicators identified thus far include:

- The ranges, distributions, and time-series of simulated flows and storage volumes in:
    - the Water Conservation Areas (WCA),
    - the Stormwater Treatment Areas (STA),
    - the Everglades National Park (ENP),
    - the Lower East Coast (LEC) service areas,
    - the Everglades Agricultural Areas (EAA), and
    - Lake Okeechobee (LOK).
- Percentage of time discharges are no greater than a specified (e.g., 2010) base flow condition in a specified (e.g., LEC) area.
- Percentage time seaward groundwater flow at coastline of a specified (e.g., LEC) area.
- Frequency, severity, and duration of water shortages compared to a specified (say 2010) base condition in a specified (e.g., LEC) area.
- Number of months under supply-side management in a specified (e.g., LOK) area.
- Percentage of supplemental demands met when compared to a (e.g., 2010) base case in a specified(e.g., LOK) area.

- Reduction in agricultural crop losses compared to those of a (e.g., 2010) base condition in a specified (e.g., LEC) area.
- Reduction in landscape and convenience losses compared to a (e.g. 2010) base condition in a specified (e.g., LEC) area.
- Total and average cost of additional water supplies in a specified (e.g., LEC) area.
- Percent of time lake stages exceed a specified elevation (e.g., 15.0 feet msl) for a specified (e.g., 26-year) period of record for specified (e.g., 1990 and 2010) base cases in a specified (e.g., LOK) area.
- Percent of time water levels within littoral zone fall below a specified elevation (e.g., 13 feet) compared against specified (e.g., 1990 and 2010) base conditions in a specified (e.g., LOK) area.
- Frequency, duration and extent of violating maximum and minimum desired ranges of flows to specified areas (e.g., the Lt. Lucie Estuary and the Caloosahatchee Estuary) for specified (e.g., 1990 and 2010) base case in a specified (e.g., LOK) area.
- Frequency, duration and extent of deviations from natural flow conditions for specified (e.g., 1990 and 2010) base cases to specified regions (e.g., the Shark Slough Basin) from other specified regions (e.g., the WCA-3 of the ENP area).
- Volume and timing of distribution of water delivered to various sites during a specified (e.g., 26-year) period of record as compared to the NSM or to a specified (e.g., 2010) base case in a specified (e.g., ENP) area.

Relative Rankings. At the workshop an attempt was made to rank the various sources of risk and uncertainty as to their relative impact on various system performance indicators. This proved difficult because the various parameters and inputs interact and have different effects at different times depending on the values of other variables. Furthermore, the grid-cell locations of the uncertain parameters with respect to the grid-cell locations of a particular system performance indicators are also important. We concluded from the workshop discussions that none of the above sources of risk and uncertainty could be eliminated from consideration; each may have a significant impact on at least one system performance indicator at one or more grid-cell locations.

Sensitivity Analysis

The usefulness of any hydrologic simulation model is dependent on the accuracy and reliability of its output data. Yet, because all simulation models are imperfect abstractions of reality, and because accurate input data are rarely available, all output values are subject to errors. The modeling errors and input data errors are not independent of each other. They can interact in various ways. The end result are errors in predictions and uncertainty associated with model output.

A number of sensitivity analysis techniques have been developed to track and account for errors in a simulation and to characterize the resulting ranges of uncertainties. Most approaches examine single sources of error relative to some fixed (most likely) conditions with regard to all other potential sources of error, and determine the sensitivity of model output values to changes in that single source of error. Sensitivity analyses can be extended to examine the combined effects of multiple sources of error, as well. This report will focus on those techniques of sensitivity analysis identified during the workshop that seem to be most applicable to the hydrologic modeling activities of the District.

Modeling and Input Data Errors

Before describing some practical sensitivity analysis techniques, it may be useful to consider the general subject of modeling errors and input data errors. These input data errors include those associated with parameter values, operating policies and boundary conditions.

Modeling Errors. Errors due to lack of understanding and lack of sufficient theory, regarding the system and hydrologic processes being modeled, i.e., errors due to modeling a simplified version of reality, will always exist. A model can always be modified to better fit one's understanding of reality by increasing its complexity. However, doing this will not always improve the precision of the model output, particularly if the increased complexity is based on processes whose parameters are more difficult to measure. A number of studies have addressed model simplification, but no one has been able to identify that level of model detail needed to minimize modeling related errors in model output values.

The problem of determining the "optimal" level of modeling detail is particularly important when simulating the hydrologic events at many sites over large areas. Perhaps the best approach for these types of simulations is to establish confidence levels for alternative sets of models and then statistically compare simulation results. But even this is not a trivial or costless task. Increases or decreases in the temporal or spatial resolution of SFWMM or NSM would require considerable data collection and/or processing, model recalibrations, and possibly the solution of stability problems resulting from the numerical methods used in the models. Obtaining and implementing alternative hydrologic simulation models (such as Mike-11 or Mike-SHE from the Danish Hydraulics Institute) would involve considerable investments of money and time for data preparation and for learning how to use, maintain, and modify the models, as required.

Input Data Errors. Errors and uncertainties associated with particular input data at a particular time period and at a particular site (grid cell) will have different affects on model output variable values, and derived system performance indicators. These affects will depend on what the output values or performance indicators are, what time period they apply, and where they are located, relative to the time and place of the particular input data. It is generally true that only a relatively few input variables dominate or substantially influence the values of a particular output variable or performance indicator at a particular location (grid-cell site) and time. However, with many output variables and performance indicators being considered, and at many sites and in many time periods, practically all input data errors must be considered. For large scale hydrologic simulation models such as SFWMM and NSM, it is not clear which if any of the input data uncertainties one can ignore when performing sensitivity analyses on all model output data. If the range of uncertainty of only some of the output data are of interest, then undoubtedly only those input data whose uncertainty has a significant impact on the values of those output data need be included in the sensitivity analysis.

If input data estimates are based on repeated measurements, a frequency distribution can be estimated that characterizes natural variability. The shorter the record of measurements, the greater the uncertainty regarding the long-term statistical characteristics of that variability. If obtaining a sufficient number of replicate measurements is not possible, subjective estimates of input data ranges and probability distributions are often made. Using a mixture of subjective estimates and actual measurements does not affect the application of various sensitivity analysis methods that can use these sets or distributions of input values, but it may affect the conclusions that can be drawn from the results of these analyses.

It would be nice to have available accurate and easy to use analytical methods for relating errors in input data to errors in model outputs, and to errors in system performance indicator values that are derived from model output. Such analytical methods do not exist for complex simulation models like SFWMM and NSM. However methods based on simplifying assumptions and approximations can be used to yield useful sensitivity information.

One measure of sensitivity is the sensitivity coefficient. This is the derivative of a model output variable with respect to an input variable or parameter. A number of sensitivity analysis methods use these coefficients. First-order and approximate first-order uncertainty analyses are two such methods that will be discussed later in this section. The difficulty of obtaining the derivatives for many models, the need to assume linear relationships (when obtaining estimates of derivatives by making small changes of input data values near their nominal or most likely values), and the large variance associated with most hydrologic process models, have motivated the replacement of analytical methods by numerical and statistical approaches to uncertainty analysis. Monte Carlo sampling of input data distributions has been one of the most commonly used of these statistical approaches.

Users of Sensitivity Analyses

"Sensitivity analysis" is a term used to describe the investigation of the importance of error or of uncertainty in various parameters or inputs to a decision or modelling process. The exact character of that analysis depends upon the particular context and the questions of concern. Sensitivity studies can provide a general assessment of model precision when used to assess system performance for alternative scenarios, as well as detailed information addressing the relative significance of errors in various parameters. As a result, sensitivity results can be of interest to the general public, federal and state management agencies, upper level District administrators and the Board of Directors, model users within the District, and model developers. Clearly, upper District management and the public may be interested in more general statements of model precision, and should be provided such information along with model predictions. On the other hand, detailed studies addressing the significance and interactions among individual parameters would only be meaningful to model developers and some model users. Only they can use such data to interpret model results and to identify where efforts to improve models should be directed.

A consensus at the workshop was that initial sensitivity analysis studies would focus on two products:

(1) detailed results to guide research and assist model improvement efforts, and

(2) calculation of general descriptions of uncertainty associated with model predictions so that policy decisions can reflect both the modelling efforts best prediction of system performance and the precision of such predictions.

In the first case, the target audiences would include model developers, model users, and those responsible for directing and funding model improvement and data collection efforts related to modelling studies. These communities should profit from a clear methodology for investigating and presenting the relative uncertainty in model projections due to possible errors in different sets of parameters and input data. The resulting sensitivity analyses should assist in the direction of efforts to improve the precision of model projections. They should also contribute to a better understanding of the relationships between model assumptions, parameters, data and model predictions.

For the second case, the target audience is very broadly all those interested in model prediction. This would include those interested in development of policies, review of policies, or just predictions of future system performance. The relative precision associated with different model prediction should have a significant effect on policy development. For example, the analysis may show that, given data inadequacies, there are very large error bands associated with some flows and stages in the Everglades under natural conditions. When such large uncertainties exist, it is appropriate that results be used with appropriate skepticism and that incremental strategies be explored so that greater experience with the operation of new facilities can accumulate.

Sensitivity Analysis Philosophies

In the field of operations research, sensitivity analysis features are available in many linear and nonlinear programming packages. They allow the evaluation of the effect on the objective function of a change in the resources available to the system, and a change in levels set for various constraints. Thus sensitivity analysis addresses the change in "optimal" system performance associated with changes in various parameter values, and also how "optimal" decisions would change with changes in resource constraint levels, or target output requirements. This kind of sensitivity analysis allows the analyst to ask how much another unit of resource would be worth, or what "cost" a proposed change in a constraint places on the optimal solution. Thus they help design decisions.

Sensitivity analysis is also a term widely used in the engineering economics literature. The engineering economics literature is generally concerned with the economic implications of various decisions, and the possible impact of input parameter uncertainty on the benefits associated with those decisions. Often one must make a decision, and only later can that person determine what the values were for various parameters which affected the economic outcome of the project. Thus in the engineering economic literature, sensitivity is often explored by looking at the economic benefits (or costs) that would be achieved with various fixed decisions if alternative sets of parameter values are realized.

Here the emphasis is on evaluating decisions, and the cost of mistakes. The question then becomes: which decision should be selected based upon a matrix of values? Even when parameter errors result in decisions very different from the optimal one, there may still be little loss of economic benefits (Loucks et al., 1981, pp. 124-129). This is often true when there will be opportunities to correct mistakes (phased expansion). One often finds that while the best or optimal value of a decision variable may be sensitive to the values of several parameters, the actual economic benefits associated with each decision may not be. However, errors in estimated costs could be high even though the loss of economic efficiency from a non-optimal decision was small.

The next section develops a simple sensitivity analysis procedure. It is based on the idea of varying one uncertain parameter value, or set of parameter values, at a time. This approach is very similar those most often employed in the engineering economics literature. The ideas are applied to a water quality example to illustrate their use. Subsequent sections develop first-order uncertainty analysis and Monte Carlo sampling techniques. The latter employ probabilistic descriptions of the possible errors in different sets of parameter values. However, the first-order method is simpler and less computational demanding because it uses a first-order linear approximation to changes in model output with changes in the various parameter values. The Monte Carlo method provides a fuller and richer description of the impact of parameter errors on the resulting errors in model output and system performance indices.

Deterministic Sensitivity Analysis

A deterministic sensitivity analysis approach is presented here that is similar to the sensitivity analysis procedures presented in the engineering economics literature. The primary difference is that the output variable of interest is not assumed to be expressed in monetary terms nor is an economic performance index to be maximized. Thus one does not know if more or less of a given variable is better or worse. Perhaps too much and/or too little is undesirable. The key idea is that, whether employing physical measures or economic metrics of performance, various parameters (or sets of associated parameters) are assigned high and low values. Such ranges may reflect either minimum and maximum values for each parameter, the 5 and 95 percentiles of a parameters distribution, or points corresponding to some other criteria. The system model is then run with the various alternatives, one at a time, to evaluate the impact of those errors in various sets of parameter values on the output variable. Table 1 illustrates the character of the results that one would obtain. Here Y_0 is the nominal value of the model output when all parameters assume the estimated best values, and $Y_{i,L}$ and $Y_{i,H}$ are the values obtained by increasing or decreasing the values of the i^{th} set of parameters.

A simple water quality example is employed to illustrate this deterministic approach to sensitivity analysis. The analysis techniques illustrated here are just as applicable to complex models, such as SFWMM and NSM. The primary difference is that more work would be required to evaluate the various alternatives with a more complex model, and the model responses might be more complicated.

Table 1.
Sensitivity of Model Output Y to Possible Errors in Four Parameter Sets

Parameters adjusted°	Low value	Nominal	High Value
Parameter-set #1:	$Y_{1,L}$	Y_0	$Y_{1,H}$
Parameter-set #2:	$Y_{2,L}$	Y_0	$Y_{2,H}$
Parameter-set #3:	$Y_{3,L}$	Y_0	$Y_{3,H}$
Parameter-set #4:	$Y_{4,L}$	Y_0	$Y_{4,H}$

°A "parameter set" is a single parameter or a group that vary together.

A simple water quality model is provided by Vollenweider's empirical relationship for the average phosphorus concentration in lakes (Vollenweider, 1976). He found that the phosphorus concentration, P, is a function of the annual phosphorus loading rate, L, the annual hydraulic loading, q, and the mean water depth, z.

$$P = (L/q) / [\, 1 + (z/q)^{0.5} \,]$$

where

P is phosphorus concentration in lake water (mg/m^3)
L is annual phosphorus loading ($mg/m^2 \cdot a$)
q is annual hydraulic loading (m/a or more exactly $m^3/m^2 \cdot a$)
z is mean water depth (m).

L/q and P have the same units; the denominator is an empirical factor that compensates for nutrient recycling and elimination within the aquatic lake environment.

Data for Lake Ontario would suggest that reasonable values of the parameters are L = 680 mg/m^3; q = 10.6 m/a; and z = 84 m, yielding P = 16.8 mg/m^3. Values of phosphorus concentrations less than 10 mg/m^3 are considered oligotrophic, whereas values greater than 20 mg/m^3 generally correspond to eutrophic conditions. Reasonable ranges reflecting possible errors in the three parameters yield the values in Table 2.

Table 2.
Sensitivity of Estimates of Phosphorus Concentration (mg/m^3) to Model Parameter Values

	Parameter Value		*Phosphorus Conc.*	
	Low	High	P(low)	P(High)
L: P Loading ($mg/m^3 \cdot a$)	500	900	12.4	22.3
q: Hydraulic Loading (m/a)	8	13.5	20.0	14.4
z: Mean Depth (m)	81	87	17.0	16.6

One may want to display these results so they can be readily visualized and understood. A **tornado diagram** (Eschenbach, 1992) would show the lower and upper values of P obtained from variation of each parameter, with the parameter with the widest limits first, and the smallest limits last. Tornado diagrams are easy to construct and can include a large number of parameters without becoming crowded.

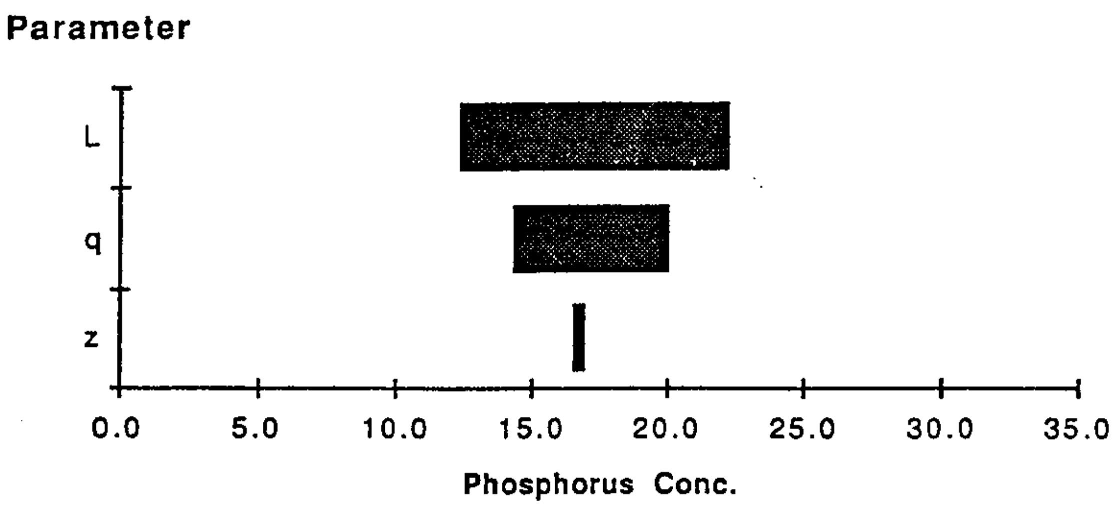


Figure 4. A Tornado diagram showing the range of the output variable representing phosphorus concentrations for high and low values of each of the parameter sets. Parameters are sorted so that the largest range is on top, and the smallest on the bottom.

Particularly when the intervals are asymmetric, tornado diagrams can be a little confusing. An alternative is a **Pareto chart** showing the width of the uncertainty range associated with each variable, ordered from largest to smallest. A Pareto chart is illustrated in Figure 5.

An alternative tool for visual presentation is a **spider plot** showing the impact of uncertainty in each parameter on the variable in question, all on the same graph (Eschenback, 1992; DeGarmo, 193, p. 401). A spider plot shows the particular functional response of the output to each parameter on a common scale, so one needs a common metric to represent changes in all of the parameters. Here we use percentage change from the nominal or best values.

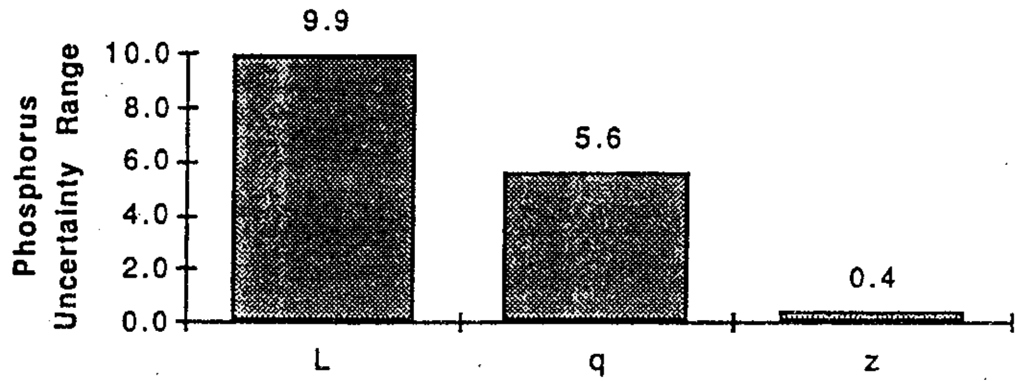


Figure 5. A Pareto Chart showing the range of the output variable representing phosphorus concentrations resulting from high and low values of each parameter set considered.

Spider plots are a little harder to construct than tornado diagrams, and can generally include only 4-5 variables without becoming crowded. However, they provide a more complete view of the relationships between each parameter and the performance measure. In particular, a spider plots reveals nonlinear relationships and the relative sensitivity of the performance measure to (percentage) changes in each variable.

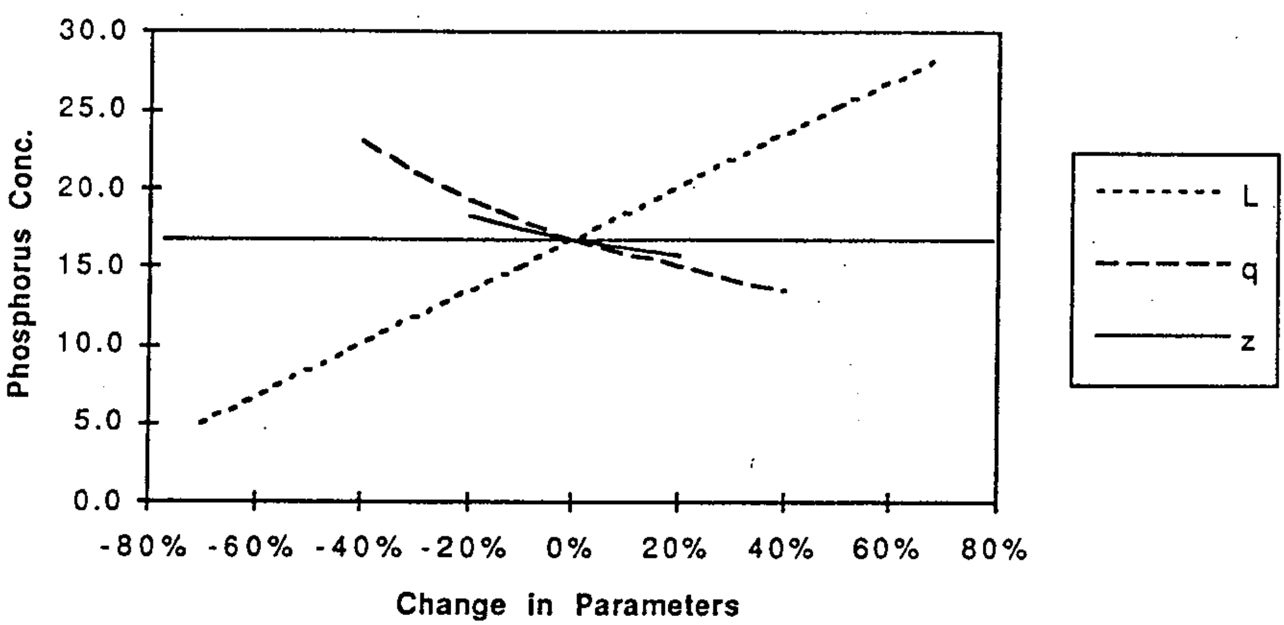


Figure 6. Spider Plot illustrates the relationships between model output describing phosphorus concentrations and variations in each of the parameter sets, expressed as a percentage deviation from their nominal values.

In the spider plot above, the linear relationship between P and L is shown, while the gentle nonlinear relationship between P and q is also illustrated. The range for z has been kept small given the limited uncertainty associated with that parameter.

Multiple Errors and Interactions. An important issue that should not be ignored is the impact of simultaneous errors in more than one parameter. Probabilistic methods directly address the occurrence of simultaneous errors, but the correct joint distribution needs to be employed. With simple sensitivity analysis procedures, errors in parameters are generally investigated one at a time, or in groups. The idea of considering pairs or sets of parameters is discussed here.

Groups of factors. It is often the case that reasonable error scenarios would have several parameters changing together. For this reason, the alternatives have been called parameter sets. For example, possible errors in water depth would be accompanied with corresponding variations in aquatic vegetation and chemical parameters. Likewise, alternatives related to changes in model structures might be accompanied with variations in several parameters. In other cases, there may be no causal relationship among possible errors (such as model structure versus inflows at the boundary of the modelled region), but they might still interact to effect the precision of model predictions.

Combinations. If one or more non-grouped parameters interact in significant ways, then combinations of one or more errors should be investigated. However, one immediately runs into a complexity problem. If each of m parameters can have 3 values (high, nominal, and low) there are 3^m combinations, as opposed to $2m + 1$ if each parameter is varied separately. [For $m = 5$, the differences are $3^5 = 243$ versus $2(5)+1 = 11$.] These numbers can be reduced by considering instead only combinations of extremes so that only $2^m + 1$ cases need be considered [$2^5 + 1 = 33$], which is a more manageable number. However, all of the parameters would be at one extreme or the other, so the cases would be very unusual.

Two factors at a time. A compromise is to consider all pairs of two factors at a time. There are $m(m-1)/2$ possible pairs for which we could consider 4 combinations of high and low values. This results in $2m(m-1)+1$ cases. [For $m = 5$ this yields 41 cases.] The presentation of these results could be simplified by displaying for each case only the maximum error, which would result in $m(m-1)/2$ cases that might be displayed in a Pareto diagram. This would allow identification of those combinations of two factors that might yield the largest errors and thus are of most concern.

For the water quality example, if one considers all pairs of factors, and all four combinations of high (+) and low (-) errors, they obtain as the absolute value of the error for each case. These are shown in Figure 7.

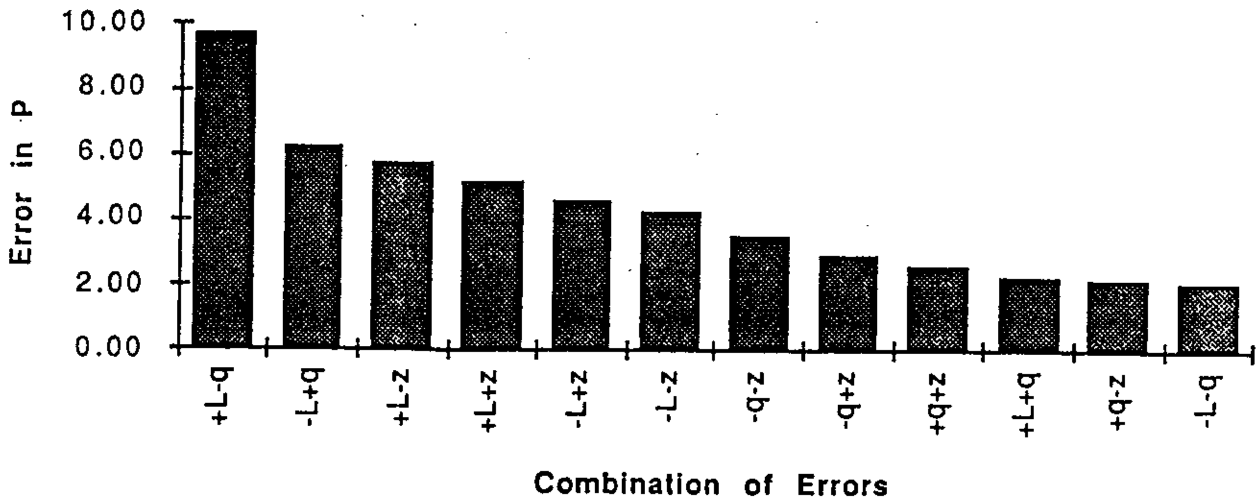


Figure 7. Pareto diagram showing errors in Phosphorus concentrations for all combinations of pairs of input parameters errors. A + indicates a high value, and a - indicates a low value for indicated parameter. L is the phosphorus loading rate, q is the hydraulic loading, and z is the mean lake depth.

Considering only the worst error for each pair of variables yields:

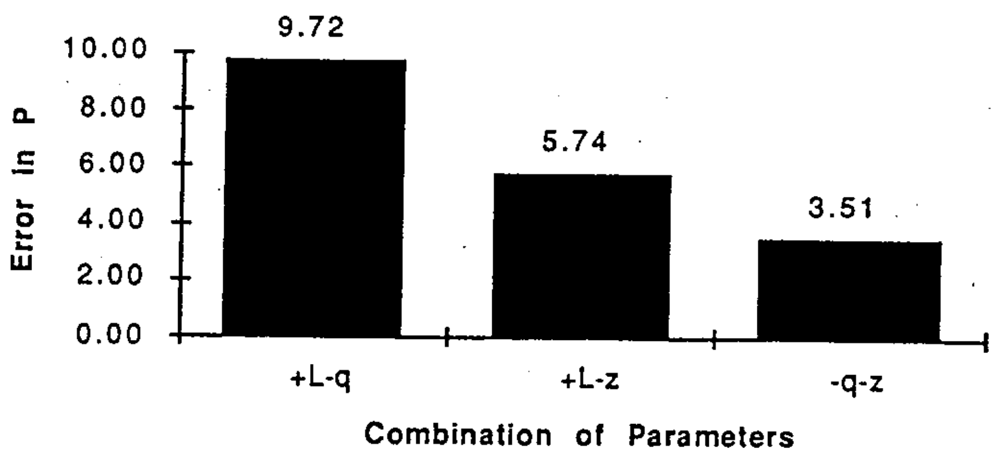


Figure 8. Pareto diagram showing worst error combinations for each pair of input parameters. A + indicates a high value, and a - indicates a low value for indicated parameter.

Here we see, as is no surprise, that the worst error results from the most unfavorable combination of L and q values. If both parameters have their most unfavorable values, the predicted phosphorus concentration would be 27 mg/m^3.

Looking for non-linearities. One might also display in a Pareto diagram the maximum error for each pair as a percentage of the sum of the absolute values of the maximum error from each factor separately. If the model of the system and the physical measure or economic metric were strictly linear, then the individual errors should add. The ratio of the joint error to the individual errors would illustrate if potentially important nonlinear interactions occur.

First-Order Uncertainty Analysis

The above deterministic analysis has trouble representing reasonable combinations of errors in several parameter sets. If the errors are independent, it is highly unlikely that any two sets would actually be at their extreme ranges at the same time. By defining probability distributions of the values of the various parameter sets, and specifying their joint distributions, a probabilistic error analysis can be conducted. In particular, for a given performance indicator, one can use multivariate linear analyses to evaluate the approximate impact on the performance indices of uncertainty in various parameters. As shown below, the impact depends upon the square of the sensitivity coefficients (partial derivatives) and the variances and covariances of the parameter sets.

For a performance indicator $I = F(\mathbf{Y})$, which is a function $F(\bullet)$ of model outputs $\mathbf{Y}$, which are in turn a function $g(\mathbf{P})$ of input parameters $\mathbf{P}$, one can use a multivariate Taylor series approximation of F to obtain the expected value and variance of the indicator:

$$E[I] = F(\text{ mean values of input parameters}) + (1/2) \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} [\partial F^2/\partial P_i \partial P_j] \operatorname{Cov}[P_i, P_j] \right\}$$

and

$$\operatorname{Var}[I] = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\partial F}{\partial P_i}\right)\left(\frac{\partial F}{\partial P_j}\right) \operatorname{Cov}[P_i, P_j]$$

where $(\partial F/\partial P_i)$ are the partial derivative of the function F with respect to P_i evaluated at the mean values of all input parameters, and $\partial^2 F/\partial P_i{}^2$ are the second partial derivatives. If all the parameters are independent of each other, and the second-order terms in the expression for the mean E[I] are neglected, one obtains

$$E[I] = F(\text{mean values of input parameters})$$

and

$$\mathrm{Var}\,[I] = \sum_{i=1}^{n} [\frac{\partial F}{\partial P_i}]^2 \,\mathrm{Var}\,[P_i]$$

(Benjamin and Cornell, 1970). The first equations for E[I] show that in the presence of substantial uncertainty, the mean of the output from nonlinear systems is not simply the system output corresponding to the mean of the parameters (Gaven and Burges, 1981, p. 1523).

It is the approximation for the variance Var[I] that interests us in the analysis of uncertainty. The contribution of P_i to the variance of I equals Var[P_i] times $[\partial F/\partial P_i]^2$, which are the squares of the sensitivity coefficients for I with respect to each input parameter P_i.

Approximate First-Order Uncertainty Analysis

It may appear that first-order analysis is difficult because the partial derivatives of the performance indicator I are needed with respect to the various parameters. However, reasonable approximations of these sensitivity coefficients are easily obtained from the simple sensitivity analysis described in Table 3, as shown below. In Table 3, three different parameter sets, P_i, are defined in which one parameter of the set is set at its high value, P_{iH}, and one at its low value , P_{iL}, to produce corresponding values (called high, I_{iH}, and low, I_{iL}) of the system performance indicator I.

Table 3.

Approximate parameter sensitivity coefficients.

Parameter Set, P_i	Low Value	High Value	Sensitivity Coefficient
Parameter set 1:	I_{1L}	I_{1H}	$[I_{1H}-I_{1L}]/[P_{1H}-P_{1L}]$
Parameter set 2:	I_{2L}	I_{2H}	$[I_{2H}-I_{2L}]/[P_{2H}-P_{2L}]$
Parameter set 3:	I_{3L}	I_{3H}	$[I_{3H}-I_{3L}]/[P_{3H}-P_{3L}]$

It is then necessary to estimate some representation of the variances of the various parameters with some consistent procedure. For a normal distribution, the distance between the 5 and 95 percentiles is 2(1.645) = 3.3 standard deviations Thus, if the high/low range is thought of as approximately a 5-95 percentile range for a normally distributed variate, a reasonable approximation of the variance might be

$$\mathrm{Var}[P_i] = \{ [P_{iH}-P_{iL}]/3.3 \}^2.$$

This is all that is needed. Use of these average sensitivity coefficients is very reasonable for modeling the behavior of the system performance indicator I over the indicated ranges.

As an illustration of the method of first-order uncertainty analysis, consider the lake quality problem described above. The "system performance indicator" in this case is the model output, the phosphorus concentration P, and the input parameters, now denoted as X = L, q, and z. The standard deviation of each parameter is assumed to be the specified range divided by 3.3. Average sensitivity coefficients $\partial P/\partial X$ were obtained using the suggested formula. The results are reported in the table below.

Table 4.

Calculation of approximate parameter sensitivity coefficients. P= Phosphorous
X= Parameter

Variable X	Units	$\partial P/\partial X$	St Dev X	$(\partial P/\partial X)^2$ Var(X)	%
L	$mg/m^2 \cdot a$	0.025	121.21	8.98	75
q	m/a	−1.024	1.67	2.91	24
z	m	−0.074	1.82	0.02	1

Assuming the parameter errors are independent:

$$\text{Var}[P] = 8.98 + 2.91 + 0.02 = 11.91$$

The square root of 11.91 is the standard deviation equal to 3.45. This agrees well with a Monte Carlo analysis reported below.

Note that 100*(8.98 / 11.91), or 75% of the total parameter error variance in the phosphorus concentration P is associated in the phosphorus loading rate L, 24% is associated with the hydrologic loading q , and only 1% with uncertainty in the mean depth z. If the uncertainty in z were eliminated, it would really have no impact on the overall model error. Likewise, reducing the error in q would at best have a modest impact on the total error.

Due to model error, the estimated phosphorus lake concentration has a standard deviation of 3.45. Thus a 5-95 percentile interval would be about

$$16.8 \pm 1.645\ (3.45)\ mg/m^3 = 16.8 \pm 5.7\ mg/m^3 = 11.1 - 22.5\ mg/m^3.$$

assuming the errors are normally distributed, so that ± 1.645 standard deviations around the mean yields a 5-95 percentile region. These error bars indicate there is substantial uncertainty associated with the phosphorus concentration P, primarily due to uncertainty in the loading rate L.

The upper bound of 22.6 mg/m^3 is substantially less than the 27 mg/m^3 that would be obtained if both L and q had their most unfavorable values. In a probabilistic analysis with independent errors, such an unlucky combination is highly unlikely.

Warning on accuracy. First-order uncertainty analysis is indeed an approximate method based upon a linearization of the response function represented by the full simulation model. It may provide inaccurate estimates of the variance of the response variable for nonlinear systems with large uncertainty in the parameters. (Gaven and Burges, 1981, provide an example; Kuczera, 1988, suggests a nonlinearity measure for check the validity of the multivariate linear approximation). In such cases Monte Carlo simulation (discussed below) or the use of higher-order approximation may be required. Beck (1987, p. 1426) cites studies that found that Monte Carlo and first-order variances were not appreciably different, and a few studies that found specific differences. Differences are likely to arise when the distributions used for the parameters are bimodal (or otherwise unusual), or some rejection algorithm is used in the Monte Carlo analysis to exclude some parameter combinations. Such errors can result in a distortion in the ranking of predominant sources of uncertainty (Beck, 1987, p. 1428). However, in most cases very similar results were obtained.

Expected value of additional information

One can also consider statistics such as:

- the expected value of sample information, and
- the expected value of perfect information.

For example, the expected value of perfect information might be measured by the variance of a performance index with natural variability and parameter error, to its variance with natural variability alone. Or, the expected value of perfect information in a single variable could be measured by the variance of a performance index due to the possible errors in all of the parameters, to its variance with the uncertainty in one particular parameter eliminated.

One might ask: if I could eliminate all parameter uncertainty so that in that sense I could select the optimal decision, what improvement in benefits or reduction in uncertainty would I obtained over the best I could expect to do on average given that I have parameter and model uncertainty? The expected value of perfect information indicates how much on average parameter uncertainty costs in terms of parameter variability, or some economic metric. If parameter uncertainty on average is costing a lot, it may pay to invest in studies or better models to eliminate it. If parameter uncertainty has a very modest impact on the expected benefits and decision, one should find other things to worry about.

If it appears that it may pay to invest to reduce parameter uncertainty, then one needs to consider how effective particular investigations would be. If one thinks of such investigations as obtaining sample information, then it is clear that the expected value of sample information should describe the increase in the expected benefits, or the reduction in the variance, that one expects to obtain from such information due to the corresponding reduction in parameter error. Some investigations may be effective at getting information related to significant sources of system uncertainty, while others are either less effective at reducing parameter errors, or are related to parameter errors that are not particularly important anyway (Binley and Beven, 1991, p. 78; Beven, 1993, p. 49). Approximate first-order uncertainty analyses can be used to see what the impact of a proposed reduction in the uncertainty of particular parameters would be on the overall uncertainty on key physical measures and economic metrics (Beck, 1987, p. 1428).

Fractional Factorial Design Method

An extension of first-order uncertainty analysis would be a more complete exploration of the response service using a careful statistical design. First consider a complete factorial design. Input data are divided into discrete "levels'. The simplest case is two levels. These two levels can be defined as a nominal value, and a high (low) value; in some discrete cases, only two parameter values may be reasonable. Simulation runs are made for all combinations of parameter levels. For n different inputs, this would require 2^n simulation runs. Hence for a three-input variable or parameter problem, 8 runs would be required. If 4 discrete levels of each input variable or parameter were allowed to provide a more reasonable description of a continuous variable, the three-input data problem would require 4^3 or 64 simulation runs. Clearly this is not a useful tool for models such as SFWMM and NSM. A fractional factorial design involves simulating only a fraction of what is required from a full factorial design method. The loss of information prevents a complete analysis of the impacts of each input variable or parameter on the output.

To illustrate the fractional factorial design method, consider the two-level with three-input variable or parameter problem. Table 5 below shows the 8 simulations required for a full factorial design method. The + and the - show the upper and lower levels of each input variable or parameter P_i where i = 1, 2, 3. If all 8 simulations were performed, seven possible effects could be estimated. These are the individual effects of the three inputs P_1, P_2, and P_3, the three two-input variable or parameter interactions, P_1xP_2, P_1xP_3, and P_2xP_3, and the one three-input variable or parameter interaction $P_1xP_2xP_3$.

Table 5.

A three-input factorial design.

Simulation Run	Value of Input P_1	P_2	P_3	Value of Output Variable Y
1	-	-	-	Y_1
2	+	-	-	Y_2
3	-	+	-	Y_3
4	+	+	-	Y_4
5	-	-	+	Y_5
6	+	-	+	Y_6
7	-	+	+	Y_7
8	+	+	+	Y_8

Consider an output variable Y, where Y_j is the value of Y in the jth simulation run. Then an estimate of the effect, denoted $\delta(Y|P_1)$, that input variable or parameter P_1 has on the output variable Y, is the average of the four separate effects of varying P_1 :

$$\delta(Y \mid P_1) = 0.25 \left[(Y_2-Y_1)+(Y_4-Y_3)+(Y_6-Y_5)+(Y_8-Y_7) \right]$$

Each difference in parentheses is the difference between a run in which P_1 is at its upper level and a run in which P_1 is at its lower level, but P_2 and P_3 are unchanged. If the effect is equal to 0, then, in this case, P_1 has no impact on the output variable Y.

Similarly the effects of P_2 and P_3, on variable Y can be estimated as:

$$\delta(Y \mid P_2) = 0.25 \{ (Y_3-Y_1)+(Y_4-Y_2)+(Y_7-Y_5)+(Y_8-Y_6) \}, \text{ and}$$

$$\delta(Y \mid P_3) = 0.25 \{ (Y_5-Y_1)+(Y_6-Y_2)+(Y_7-Y_3)+(Y_8-Y_4) \}$$

Consider next the interaction effects between P_1 and P_2. This is estimated as the average of the difference between the average P_1 effect at the upper level of P_2, and the average P_1 effect at the lower level of P_2. This is the same as the difference between the average P_2 effect at the upper level of P_1 and the average P_2 effect at the lower level of P_1 :

$$\delta(Y \mid P_1, P_2) = (1/2) \{ [(Y_8-Y_7) + (Y_4-Y_3)]/2 - [(Y_2-Y_1) + (Y_6-Y_5)]/2 \}$$

$$= (1/4) \{ [(Y_8-Y_6)+(Y_4-Y_2)] - [(Y_3-Y_1) + (Y_7-Y_5)] \}$$

Similar equations can be derived for looking at the interaction effects between P_1 and P_3, and between P_2 and P_3 and the interaction effects among all three inputs P_1, P_2, and P_3.

Now assume only half of the simulation runs were performed, perhaps runs 2,3,5 and 8 in this example. If only outputs Y_2, Y_3, Y_5, and Y_8 are available, for our example:

$$\delta(Y \mid P_3) = \delta(Y \mid P_1, P_2) = 0.5 \{ (Y_8 - Y_3) - (Y_2 - Y_5) \}$$

The separate effects of P_3 and of P_1xP_2 are not available from the output. This is the loss in information that results from fractional instead of complete factorial design.

Monte Carlo Sampling Method

The Monte Carlo method of performing sensitivity analyses, illustrated in Figure 9, first selects a random set of input data values drawn from their individual probability distributions. These values are then used in the simulation model to obtain some model output variable values. This process is repeated many times, each time making sure the model calibration is valid for the input data values chosen. The end result is a probability distribution of model output variables and system performance indices that results from variations and possible errors in all of the input values.

Monte Carlo Sampling and Simulation

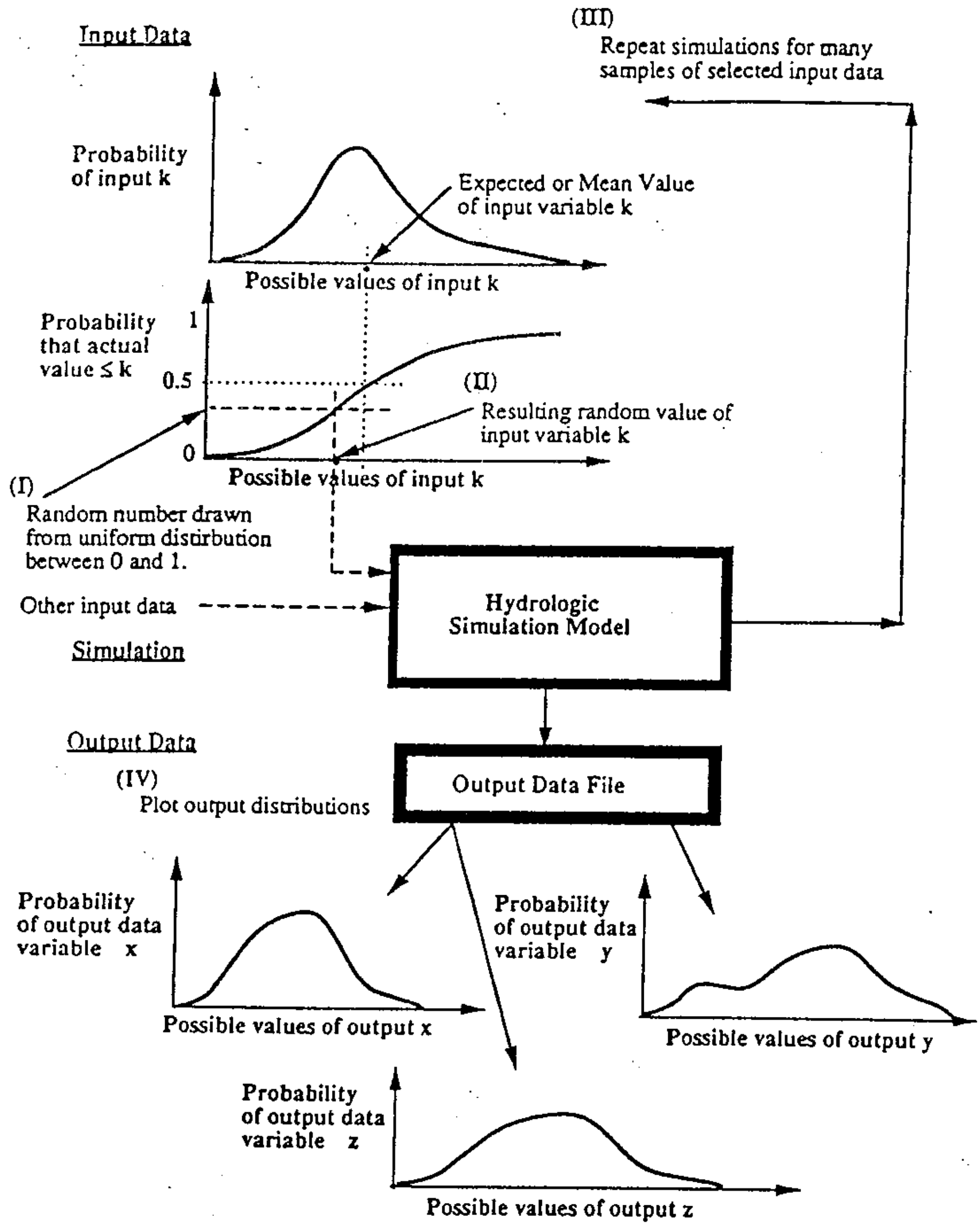


Figure 9. Monte Carlo sampling and simulation procedure for finding distributions of output variable values based on distributions, for specified confidence levels, of input data values. This technique can be applied to only one uncertain input variable at a time or to groups of input variables. The output distributions will reflect the combined effects of this input uncertainty over the specified ranges.

With a simple Monte Carlo analysis values of all of the parameter sets are selected randomly from distributions describing the individual and joint uncertainty in each, and then the modelled system is simulated to obtain estimates of the necessary performance indices. This must be done many times (often well over 100) to obtain a statistical description of system performance variability. The number of replications needed is generally not dependent on the number of parameters whose errors are analyzed. One can include in the simulation the uncertainty in parameters as well as natural processes. This method naturally evaluates the impact of single or multiple parameter uncertainty.

A significant problem that arises in such simulations is that some combinations of parameters result in unreasonable models. For example, model performance with calibration data sets might be inconsistent with available data sets. The calibration process places interesting constraints on different sets of parameters. Thus, such Monte Carlo experiments often contains checks which exclude combinations of sets of parameters which are unreasonable. In their Monte Carlo analysis, Harlin and Kung (1992) and Binley and Beven (1991) both reject sets of parameters which on a calibration data set yield calibration/validation criteria below a threshold. Thus the generated results are conditioned on this validity check.

Whenever sampling methods are used, one must be careful to consider any correlations among input data values. Sampling methods can handle spatial and temporal correlations that may exist among input data values, but the consideration of correlation requires defining appropriate conditional distributions. Users of these sampling methods must also insure that each randomly or systematically created set of input data is feasible with respect to model calibration. Of course, all replicate sets of input data values can be generated prior to their use in a simulation model, and hence they can be examined and analyzed to ensure that they meet feasibility criteria before they are used.

One major limitation of applying Monte Carlo methods to estimate ranges of risk and uncertainty for model output variable values, and system performance indicator values based on these output variable values, is the computing time required. To reduce the computing times required to perform sensitivity analyses using sampling methods, some tricks and as well as stratified sampling methods are available. The discussion below illustrates the idea of a simple adjust (or trick) with a "standardized" Monte Carlo analysis. The more general Latin Hypercube Sampling procedure is also discussed.

Simple Monte Carlo Sampling

To illustrate the use of Monte Carlo sampling methods we return to Vollenweider's empirical relationship for the average phosphorus concentration in lakes (Vollenweider, 1976). Two hundred values of each parameter were generated independently from normal distributions with the means and variances in Table 6 below.

The table contains the specified means and variances for the generated values of L, q and z, and also the actual values of the means and variances of the 200 generated values of L, q, z and also of the 200 corresponding output phosphorus concentrations, P, that were generated. Figure 10 displays the distribution of the generated values of P.

Table 6.
Monte Carlo Analysis of Lake Phosphorus Levels

Parameter:	L	q	z	P
	Specified means and standard deviations			
Mean	680.00	10.60	84.00	---
Stand. dev.	121.21	1.67	1.82	---
	Generated means and standard deviations			
mean	674.18	10.41	84.06	17.07
Stand. dev.	130.25	1.73	1.82	3.61

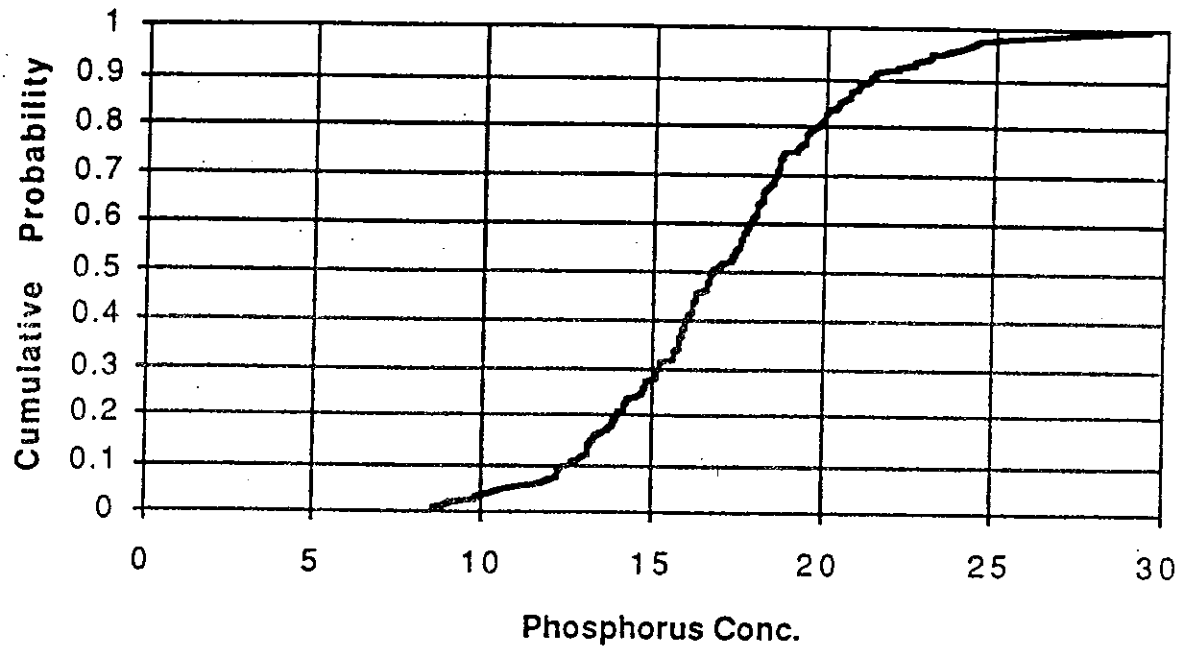


Figure 10. Distribution of lake phosphorus concentrations from Monte Carlo analysis

One can see that given the estimated levels of uncertainty that phosphorus levels could reasonably range from below 10 to above 25. The probability of generating a value greater than 20 mg/m^3 was 12.5%. The 5% to 95 percentile range was 11.1 to 23.4 mg/m^3. In the figure, the cumulative probability curve is rough because only 200 values of the phosphorus concentration were generated, but this is clearly enough to give a good impression of what the impact of the errors might be.

Sampling Uncertainty. In this example, the mean of the 200 generated values of the phosphorus concentration, P, was 17.07. However a different set of random values would have generated a different set of P values as well. Thus it is appropriate to estimate the standard error, SE, of this average. The standard error equals the standard deviation σ of the P values divided by the square root of the sample size n:

$$SE = \sigma/\sqrt{n} = 3.61/(200)^{0.5} = 0.25.$$

From the central limit theorem of mathematical statistics, the average of a large number of independent values should have very nearly a normal distribution. Thus, 95% of the time, the true mean of P should like in the interval $17.1 \pm 1.96\ (0.25)$, or 16.6 to 17.6

mg/m^3 . This level of uncertainty reflects the observed variability of P and the fact that only 200 values were generated.

Making sense of results. A significant challenge with complex models is to determine from the Monte Carlo simulation which parameters errors are important. This can be done by calculating the correlation between each generated input parameter value and the output variable value. As Table 7 below shows, based upon the magnitudes of the correlation coefficients, errors in L were most important, and those in q second in importance.

Table 7.

Correlation Analysis of Monte Carlo Results

Variable:	L	q	z	P
L	1			
q	0.079	1		
z	0.1297	–0.139	1	
P	0.851	–0.434	0.144	1

One can also use regression to develop a linear model between variations in the output and errors in the various parameters. In this case we obtain the results in the Table 8 below. The fit is very good, and $R^2 = 98\%$. If the model for P had been linear, a R^2 value of 100% should have resulted. All of the coefficients are significantly different from zero. Note that the correlation between P and z was positive in Table 8 above, but the regression coefficient for z is negative. This occurred because there is a modest negative correlation between the generated z and q values. Use of partial correlation coefficients can also correct for such spurious correlations among input parameters.

Table 8.

Results of Regression Analysis on Monte Carlo Results

	Coef.	Stand. Err.	t Ratio
Intercept	18.605	1.790	10.39
L	0.025	0.000	85.36
q	−1.068	0.022	−48.54
z	−0.085	0.021	−4.08

Finally we display a plot illustrating the relationship between P and L, where errors in L were the most important. A second plot illustrates the relation between P and z. There is considerable scatter in the P-L graph due to the generated values of the other two parameters. The plot of P versus z shows almost no relationship due to the variability introduced by the other variables. However, when those variations are controlled, the regression analysis was able to identify a statistically significant negative relationship between P and z. Such is the value of the use of regression.

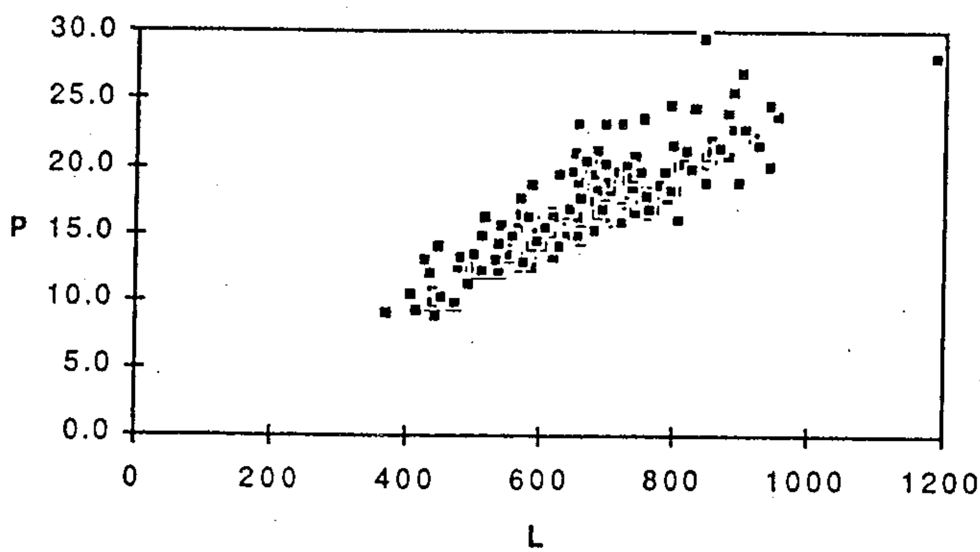


Figure 11. Distribution of lake phosphorus concentrations P versus loading L

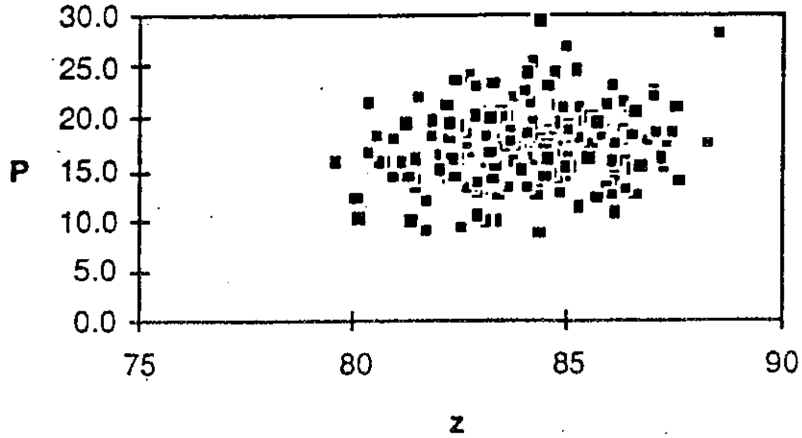


Figure 12. Distribution of lake phosphorus concentrations P versus depth z

Standardized Monte Carlo analysis

Using a "standardized" Monte Carlo analysis, one could adjust the generated values of L, q and z above so that the generated samples actually have the desired mean and variance. While making that correction, one can also shuffle their values so that the correlations among the generated values for the different parameters are near zero, as is desired. This was done for the 200 generated values to obtain the statistics shown in Table 9.

Table 9.

Standardized Monte Carlo Analysis of Lake Phosphorus Levels

Parameter:	L	q	z	P
	Specified means and standard deviations			
Mean	680.00	10.60	84.00	---
Stand. dev.	121.21	1.67	1.82	---
	Generated means and standard deviations			
Mean	680.00	10.60	84.00	17.03
Stand. dev.	121.21	1.67	1.82	3.44

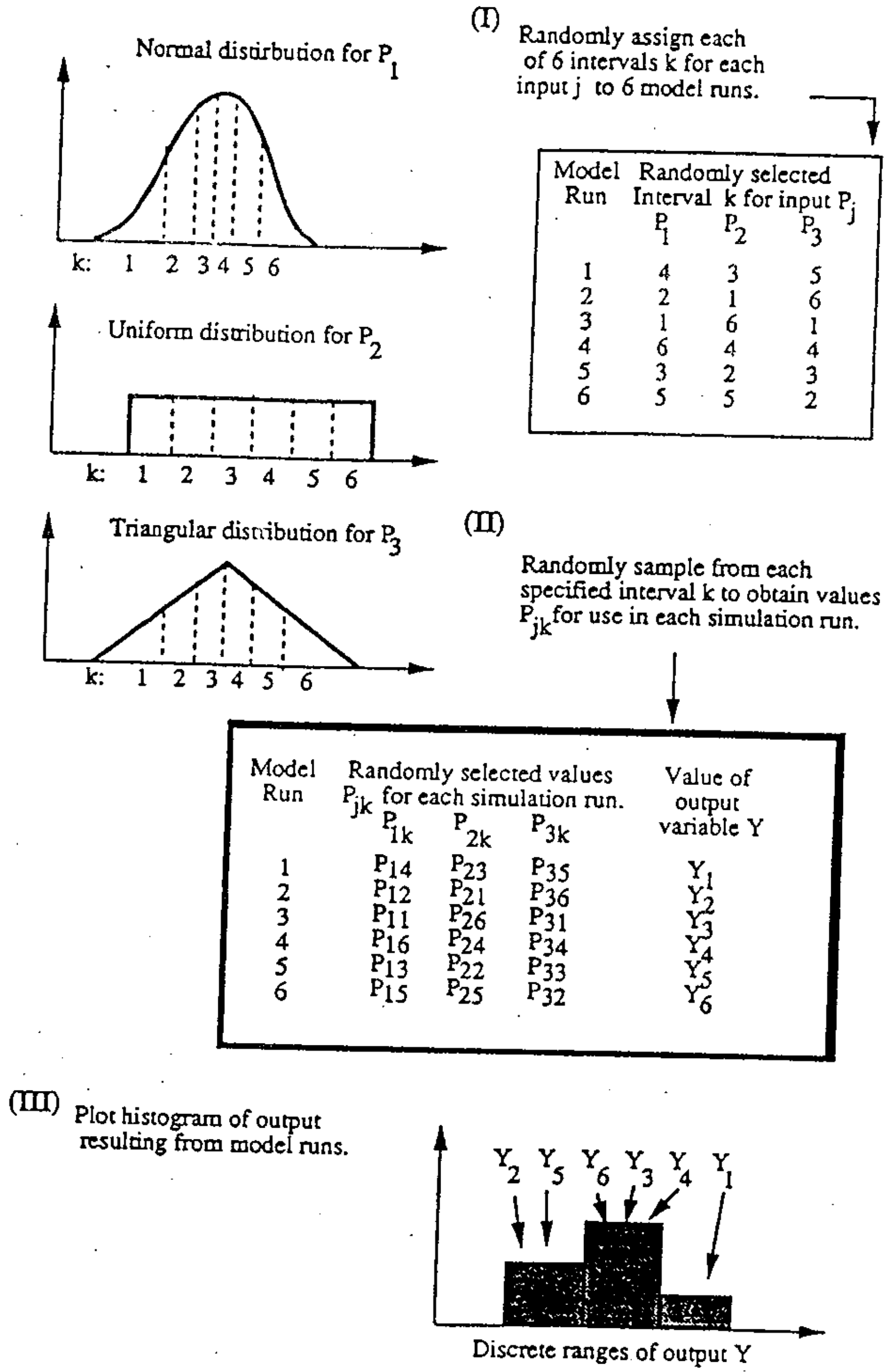


Figure 13. Schematic representation of a Latin hypercube sampling procedure for six simulation runs.

Beven (1993) and Binley and Beven (1991) suggest a Generalized Likelihood Uncertainty Estimation (GLUE) technique for assessment of parameter error uncertainty using Monte Carlo simulation. It is described as a "formal methodology for some of the subjective elements of model calibration" (Beven, 1989, p. 47). The basic idea is to begin by assigning reasonable ranges for the various parameters of a conceptual hydrologic model and then to draw parameter sets from those ranges using a uniform or some similar (and flat) distribution. These generated parameter sets are then used on a calibration data set so that unreasonable combinations can be rejected, while reasonable values are assigned a posterior probability based upon a likelihood measure which may reflect several dimensions and characteristics of model performance.

Let $L(P_i) \geq 0$ be the value of the likelihood measure assigned to the i^{th} parameter set's calibration sequence. Then the model predictions generated with parameter set/combination P_i are assigned posterior probability

$$p(P_i) = L(P_i) / \Sigma_j L(P_j)$$

These probabilities reflect the form of Bayes theorem, which is well supported by probability theory (Devore, 1991). This procedure should capture reasonably well the dependence or correlation among parameters, because reasonable sequences will all be assigned larger probabilities, whereas sequences that are unable to reproduce the system response over the calibration period will be rejected or assigned small probabilities.

However, in a rigorous probabilistic framework, the L would be the likelihood function for the calibration series for particular error distributions. (This could be checked with available goodness-of-fit procedures; for example, Kuczera, 1988.) When relatively ad hoc measures are adopted for the likelihood measure with little statistical validity, the $p(P_i)$ are best described as pseudo probabilities or "likelihood" weights.

Another concern with this method is is potential efficiency. If the parameter ranges are too wide, a large number of unreasonable or very unlikely parameter combinations will be generated. These will either be rejected or else will have small probabilities and thus little effect on the analysis. In this case the associate processing would be a waste of effort. A compromise, perhaps suggested by Beven (1993, p. 48), is to use some data to calibrate the model and to generate a prior or initial distribution for the parameters that is at least centered in the best range. Then use of a different calibration period to generate the $p(P_i)$ allows an updating of those initial probabilities to reflect the information provided by the additional calibration period with the likelihood measures adopted.

After the accepted sequences are used to generate sets of predictions, the likelihood weights would be used in the calculation of means, variances and quantiles, rather than the customary procedure of giving all the generated realizations equal weight. The resulting conditional distribution of system output reflects the initial probability distributions assigned to parameters, the rejection criteria, and the likelihood measure adopted to assign "likelihood" weights.

Calibration of Hydrologic Simulation Models

Model calibration, uncertainty estimation, and sensitivity analyses of simulation models have received considerable attention in the water quality literature because of their importance (Beck, 1987; Beven, 1993). In particular, Beven (1989; 1993) argues that the manageable representations that modelers can use to describe hydrologic systems are of necessity dimensionally-limited versions of reality with lumped descriptions of nonlinear sub-grid processes; thus uncertainty and the validity of predictions is a primary issue with water quality modeling efforts.

It was beyond the scope of the workshop to address all of the problems that have been encountered during parameter identification and model calibration. However, it is obvious that with noisy and limited data sets, and complex hydrologic and water quality models, parameter estimation is not a trivial exercise (Beck, 1987; Gupta and Sorooshian, 1985a,b; Sorooshian, et al., 1993). Thus the need for sensitivity analysis to study the consequences of uncertainty in model structures and parameter values.

Advantages of Automatic Calibration Packages

The District could employ an automatic calibration package to assist with manual trial-and-error investigations. Automatic calibration packages have three important features that can be of value. The first feature is a criterion function that must be developed for the calibration process. This criterion function is usually one involving the differences between calibrated model output values and observed or measured values. Studies have shown that greatly improved results are obtained if that criterion function reflects the character and statistical properties of the errors in the calibration data set so that appropriate statistical inferences can be made (Sorooshian et al., 1983). Data may vary dramatically in quality and errors may be cross-correlated. A better statistical description of the reliability of the input calibration data set can greatly enhance ones ability to identify good and physically reasonable data sets (Duan et al., 1988; Sorooshian et al., 1993).

The second feature of automatic calibration packages is an automated search routine that tries different parameter combinations so as to identify (within constraints set by the modeller) the combination of parameter values most consistent with the observed data set. Research in computer science has resulted in the development of several families of numerical optimization algorithms; hydrologic studies have demonstrated the value of both direct-search and derivative-based optimization algorithms for different problems (Henrickson et al., 1988). Derivative-based algorithms can be much faster, if discontinuities in various hydrologic functions do not cause too many irregularities in the response surface and its first and second derivatives. However, studies continue to show that the response surfaces often have local optima and alternative sets of parameters may result in very similar predictions with observed data sets (Sorooshian et al., 1993). These difficulties suggest the use of randomized and robust search techniques (e.g., genetic algorithms) that adequately explore alternative but physically-reasonable parameter combinations.

The third feature of automated procedures, that employ statistically reasonable objective function, is that they can generate a statistical description of the precision of the estimate parameters. This capability is explored in the next section.

Statistical Descriptions of Parameter Precision

Automated calibration routines based on a reasonable statistical description of the character of the observed data set provide statistical descriptions of the precision of parameter estimates and their covariances. Both the first-order uncertainty analysis techniques and the Monte Carlo sampling procedures need an approximation of the relative uncertainty in various parameters, and the correlations among parameters. These could be developed by some trial-and-error analyses with the simulation, or by "experts" providing their subjective judgement. Both of these procedures will no doubt be needed.

The District could also employ, where it can, statistically-oriented automatic calibration procedures which provide a statistical description of parameter precision. Such procedures can be applied to subbasins separately for the calibration of the model parameters describing that part of the system.

Often a very important contribution provided by statistically oriented parameter estimating methods (based upon the likelihood function for some assumed model error distribution) is an estimate of the covariance matrix of the estimated parameters (Kuczera, 1988; Schweppe, 1973). Because many parameters can compensate for errors in others, the uncertainty in individual parameters can overstate the true parameter uncertainty for a model as a whole given the constraints imposed by calibration data sets. In Kuczera's examples, correlations of -0.994 and -0.997 had major impacts on calculated prediction uncertainty. In Van Straten's (1983) analysis of Lake Ontario, correlations among parameters reduced prediction errors by an order of magnitude. Beck (1987, p. 1426) discusses this and other examples. The lesson for the District is that for both Monte Carlo and first-order uncertainty analyses, care should be employed when developing estimates of the variances and covariances (dependencies) of parameter estimates in the proposed sensitivity analyses the District conducts.

Sensitivity Analysis Program

The following steps outline a program that the District could follow to develop sensitivity information associated with model outputs and system performance indicators.

1. Identify parameter sets to be considered. Identify input variables and parameters whose uncertainty is to be investigated and whose impact on selected output variables is of interest. Output variables should be selected so that the time as well as space dependence of sensitivity conclusions can be assessed.

2. Specify ranges for parameter sets. Identify for each parameter, or set of parameters, a nominal (or most likely) value, and high/low or pessimistic/optimistic values at some confidence level. These high/low or pessimistic/optimistic values may be either minimum and maximum realistic quantities for each parameter, or 5 and 95 percentiles of the likely range (or some other quantiles), or points corresponding to some other criteria. To the extend possible, the intervals for each parameter should reflect the same degree of likeliness so they can be compared on an equal basis. Use of an automatic calibration package that provides a statistical assessment of the precision of parameter estimators based upon an appropriate likelihood criterion function would be most helpful. In the absence of such an aide, modelers would face the difficult task of establishing such bounds through a trial and error investigation of different magnitudes of errors in various sets of parameters. Parameter error ranges in many cases may be set by simulation/calibration to the appropriate subarea of the District.

3. Select system performance indicator. Determine one or more descriptions of system performance, economic metrics, or aggregations of physical measures of system response that will be used to investigate the impact of parameter uncertainty. Simple aggregation algorithms are

- maximum absolute deviation from nominal values over time and space
- average absolute deviation from nominal values over time and space
- maximum relative deviation from nominal values over time and space
- average relative deviation from nominal values over time and space
- importance weighted average of relative deviation from nominal values over time and space

For quality constituents that vary more than a factor of 2 or 3 over time and space, one should consider whether the relative or absolute changes should be considered. If one is interested how much a parameter affects the concentration of a certain water quality constituent, or flow levels across a large area, then it might be appropriate to look at relative deviations from the nominal values. If one is interested in dissolved oxygen deficits or sediment fluxes, then, because the big values are the ones of interest, absolute deviations from nominal values would probably be more appropriate.

It is also possible to give values at different places or different times various weights reflecting their relative importance to the issue of concern. Importance weighting is often preferred when comparing maximum, average, and weighted measures because it allows the analyst to spread weights across the areas and time periods of interest. For some systems indicators related to plant and animal species survival, particular time periods and areas may be particularly critical and should be emphasized.

One may prefer to have different aggregations for different issues. For initial screening, a straight average over time and space provides a general average assessment of the impact of uncertainty. As a companion, use of the maximum change over time and space highlights the worst local impacts which might otherwise be lost if averaged over all space and time.

4. Evaluate the impact of these ranges. The spatial and temporal impacts can be displayed using

- tables such as Table 1
- Tornado diagrams (showing lower and upper limits for each parameter) such as Figure 4.
- Pareto diagrams (showing histogram of total impact for each parameter) such as Figure 5.
- Spider plots (showing how metric changes as function of parameter changes) such as Figure 6.

Spider plots show how the aggregate index changes with variations in each of the parameters. Often, additional points are added in between the high/low or pessimistic/optimistic extremes to better represent intermediate system response. One must decide how to express parameter changes in units that can be compared. Tornado diagrams and Pareto charts rank all input variables with respect to the impact their errors are likely to have on each output variable.

5. Approximate First-Order Uncertainty Analysis. Using the results of the simple sensitivity analysis using high and low values of various parameters, estimate the corresponding sensitivity coefficients (partial derivatives with respect to changes in each parameter set) for the model using the difference method described in the text. Likewise, use the parameter ranges developed in step 2 to estimate the variance associated with each set of parameters. Use these sensitivity coefficients and variances for each parameter set to obtain an estimate of the variance of the performance measure being considered. These variances can be used to assign error bars to the reported numbers reflecting the uncertainty due to parameter error. Rank all input variables with respect to the impact their errors have on the variance of each output variable.

6. Comparison of Methods. Compare the rankings obtained using the deterministic analysis leading to results displayed in Tornado diagrams and Pareto charts with that obtained from the approximate first-order uncertainty analysis. If differences appear, determine their cause if possible. It may be that the high/low values for some variables tend to exaggerate their uncertainty, errors in some variables may interact in important ways, or nonlinearities in the model cause the error range to blow up or to be compressed. This comparison should lead to better confidence in the sensitivity analysis results, as well as a better understanding of the impacts and interactions of errors in various parameters.

7. Monte Carlo Simulation. In general, Monte Carlo simulation with the required multiple model runs is not an attractive analysis procedure for the District. In general, first-order uncertainty analyses will provided the needed guidance and estimates of precision for model predictions. However, for the District's nominal case, it may want to consider conducting a Monte Carlo simulation analysis to confirm the results on the first-order uncertainty analyses. Latin Hypercube sampling can insure that with a limited number of samples the entire domain of the errors in the individual parameters are represented in the analysis. It may also be necessary to include a check or calibration analysis step n the simulation to insure that generated combinations of parameters are reasonable.

Uncertainty in Predictions and Parameter Error. First-order uncertainty analysis and Monte Carlo sampling, as described here, both attempt to evaluate the impact of errors in parameters on the precision of predictions. In this sense, if the model calibration procedure was repeated using different data sets, another combination of sets of parameter values would result. Those values would yield different simulated system behavior, and thus different predictions. We can call this parameter error in the predictions, because it is an error that results from error in the parameters. If such parameter error were eliminated, then the prediction would always be the same and so the parameter error in the predictions would be zero. But this does not mean that predictions would be perfectly accurate.

In the calibration of water quality models, an exact fit can not be obtained because even our most complex models are only practical descriptions of the real world phenomena they describe, and our descriptions of inputs are only spatial and temporal averages of the actual inputs. No matter how good our parameter estimates, the models are not perfect and there is a residual model error. Thus, true prediction precision is a combination of model error (due to an imperfect description of the world) and parameter error (due to not having the

Carlo uncertainty analysis (as described above) should be understood to reflect only parameter error in the predictions, and do not include model error. Kuczera (1988) provides an example of a conceptual hydrologic modelling exercise with daily time steps where model error dominated parameter error.

Displaying Model Performance Indices

The most common graphical displays of system performance indicators currently used in the District include:

- Time-series plots for continuous time-dependent indicators (Figure 14a)
- Probability exceedance distributions for continuous indicators (Figure 14b),
- Histograms for discrete event indicators (Figure 14c), and
- Overlays on maps for space-dependent discrete events (Figure 14d).

Alternative Displays of System Performance

Y = Model output
F(Y) = System Performance Indicator

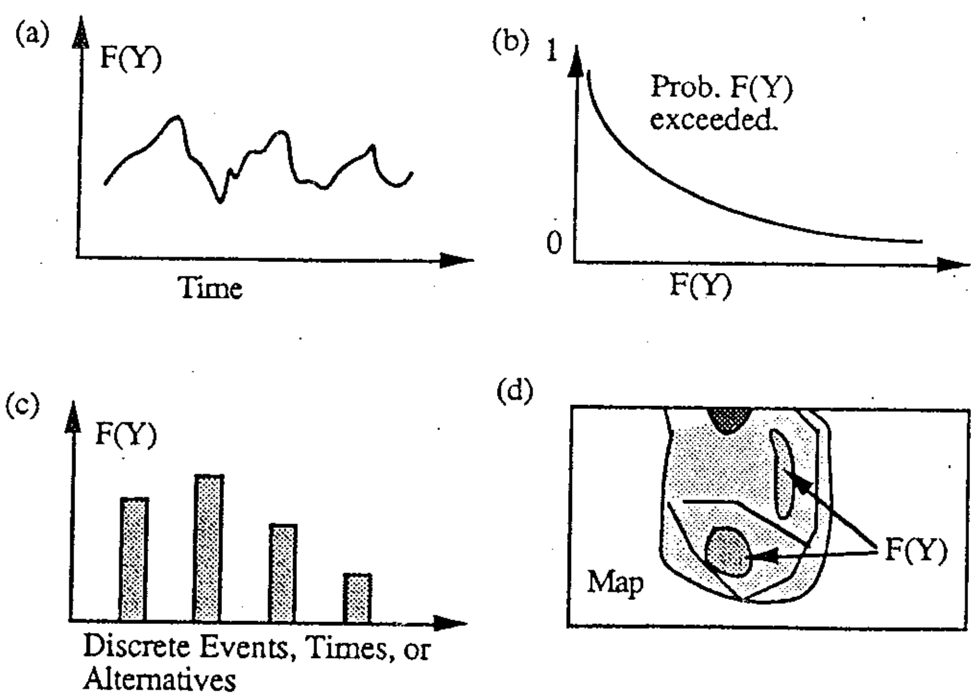


Figure 14. Different types of displays used to show model output or system performance indicator values.

As suggested in Figure 1, the first three of the above graphs could show, in addition to the single curve or bar that represents the most likely output, a range of outcomes, each associated with a given confidence interval. For overlays of information on maps, different colors could represent the spatial extents of events associated with different ranges of risk or uncertainty. Figure 15, corresponding to Figure 14, illustrates these approaches for displaying these ranges.

Alternative Displays of System Performance together with specified confidence ranges.

Y = Model output
F(Y) = System Performance Indicator

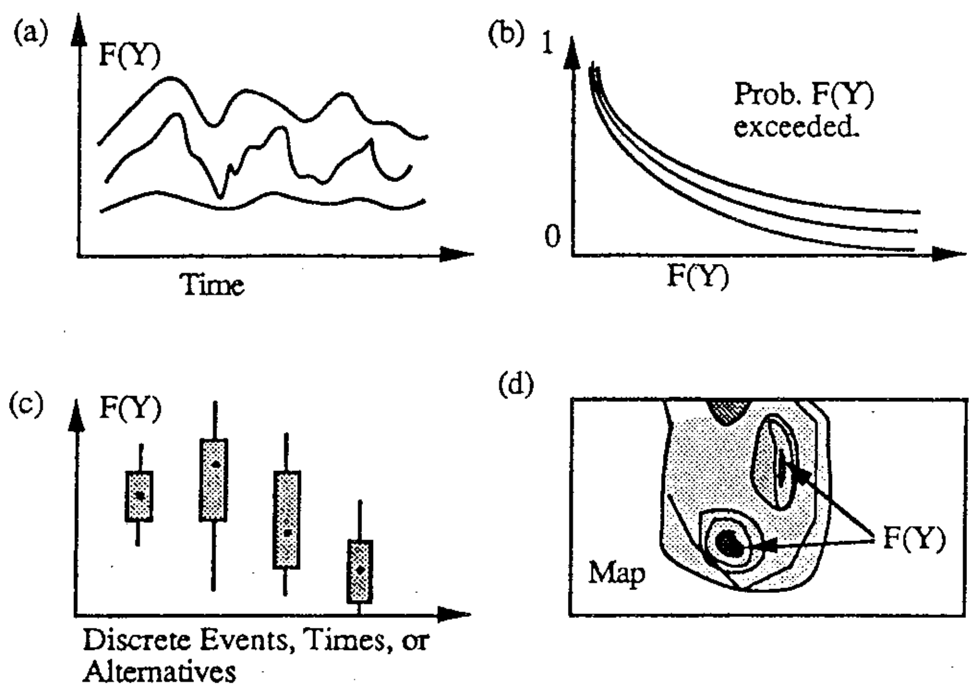


Figure 15. Plots of ranges of possible model output or system indicator values for different types of displays.

Measures to Reduce Risk and Uncertainty

A major interest of the District is increasing the precision of predictions of system operation with conditions similar to those observed historically, and also with significantly different conditions. As Figure 3 illustrates, when the conditions of interest are very different from those available to calibrate a model, prediction precision may suffer significantly.

A major purpose of sensitivity analyses is to determine which parameter and modelling errors have the largest impact on prediction uncertainty. This can be done deterministically as in Figures 4-6, or stochastically as in Table 4. Figures 4-6 illustrate the range of errors in a target system performance index that results from possible errors in different parameters. Table 4 employs a probabilistic analysis to provide an answer to the same question. Clearly, the parameters whose errors can result in large errors in system performance indices are those whose values need to be resolved better to improve the absolute precision of predictions of system performance. The sensitivity analysis program discussed above in intended to identify which parameters are likely to have errors that will have the largest impacts on model predictions of system performance.

Once the "critical" parameters have been identified, it will be the job of District personnel to develop programs to reduce the uncertainty in those parameters to the extent possible. How this can be done depends upon the parameters in question. It may require the collection of more accurate data on land surface features, or on the hydrologic characteristics of different land cover types. More accurate data can be collected on infiltration parameters, or more detailed models of different hydrologic processes can be developed. Such efforts may involve extensive field investigations.

In some case there may be little one can do to reduce some uncertainties. For example, models may be unavailable to predict the response of some bird or fish species to changes in water levels and their constituent concentrations in the Everglades, or it may not be possible to reconstruct a more accurate description of how the system was operated during a period, such as the 1950's, which is now used to calibrate the models.

As the discussion of the expected value of sample information points out, one can also use the sensitivity analysis methodologies, particularly first-order uncertainty analysis, to investigate what impact reduction of any uncertainty in specific variables will have on the precision of various system performance indices. This is a very importance issue to consider before beginning expensive field investigations.

Distinguishing Differences between Performance Indicator Distributions

Many of the studies the District will perform correspond to a comparison of the simulation output obtained with alternative sets of parameters or policies. For the purposes of discussion, suppose that the District is interested in comparing the magnitudes of flows or stages in the Everglades with two policies by simulating the system with 50 years of meteorological data. A reasonable question to ask is if observed differences are statistically significant; that is, can one really tell if the operation of the system changed, or are the observed differences explainable by random variations attributable to variations in the meteorological inputs and how the system responds.

This is a common statistical issue which is addressed by standard hypothesis tests (Devore, 1991; Benjamin and Cornell, 1970). Selection of an appropriate test requires that one first resolve what type of change one expects in the variables. Let Y_1 denote the flow magnitudes with the first policy, and Y_2 the magnitudes with the second. In many cases, one would expect policy 2 to either increase or decrease the Y levels. This can be described as a difference in the mean of the variables; for example, $E[Y_1] < E[Y_2]$. Sometimes such a change is expressed as a difference in the median (50 percentile) of the two distributions.

Alternatively, one could look for a change in the variability or variance, or a shift in both the mean and the variance. Changes described by a difference in the mean or median often make the most sense and many statistical tests are available that are sensitive to such changes. For such investigations parametric and non-parametric tests for paired and unpaired data can be employed.

Consider the differences between "paired" and "unpaired" data. The simulation results the District will examine will probably be an example of paired data. Suppose that the meteorological data for 1941-1990 is used to drive a simulation model generating data as described in the table below:

Table 11.
Possible Flow Data from a 50-year Simulation

1941	$Y_1(1)$	$Y_2(1)$
1942	$Y_1(2)$	$Y_2(2)$
1943	$Y_1(3)$	$Y_2(3)$
1944	$Y_1(4)$	$Y_2(4)$
....	....	....
1989	$Y_1(49)$	$Y_2(49)$
1990	$Y_1(50)$	$Y_2(50)$

Here there is one sample, $Y_1(1)$ through $Y_1(50)$, for policy 1, and another sample, $Y_2(1)$ through $Y_2(50)$, for policy 2. However, the two sets of observations are not independent. For example, if 1943 was a very dry year, then we would expect both $Y_1(3)$ for policy 1 in that year and $Y_2(3)$ for policy 2 to be unusually small. With such paired data, one should use a paired hypothesis test to check for differences. Paired tests are usually easier than the corresponding unpaired tests which are appropriate in other cases. (For example, if one were checking for a difference in average rainfall depth between 1941-1960, and 1961-1990, they would have two sets of independent measurements for the two periods. With such data, one should use a two-sample unpaired test.)

Paired tests are generally based on the differences between the two sets of output, $Y_1(i) - Y_2(i)$. These are viewed as a single independent sample. The question is then if the differences tend to be positive (hence Y_1 tends to be larger then Y_2), tend to be negative (hence Y_1 tends to be smaller), or if positive and negative differences are equally likely (there is no difference between Y_1 and Y_2).

Both parametric and non-parametric families of statistical test are available for paired data. The common parametric test for paired data (a one-sample t test) assumes that the mean of the differences

$$X(i) = Y_1(i) - Y_2(i)$$

are normally distributed. Then the hypothesis of no difference is rejected if the t statistic

$$T = \sqrt{n}\ \bar{x} / s$$

is sufficiently large, given the sample size n, where s^2 is the sample variance.

Alternatively, one can employ a nonparametric test and avoid the assumption that the differences X(i) are normally distributed. In such a case, one can use the Wilcoxon Signed Rank test. This nonparametric test ranks the absolute values |X(i)| of the differences. If the sum S of the ranks of the positive differences deviates sufficiently from its expected value, n(n+1)/4 (were there no difference between the two distributions), one can conclude that there is a statistically significant difference between the $Y_1(i)$ and $Y_2(i)$ series. Standard statistical texts have tables of the distribution of the sum S as a function of the sample size n, and provide a good analytical approximation for $n \geq 20$ (for example, Devore, 1991). Both the parametric t test and the nonparametric Wilcoxon Signed Rank test require that the differences between the simulated values for each year be computed.

Conclusions

This report attempts to summarize remarks made at a January 1994 workshop on risk and uncertainties in hydrologic modeling. It also provides an overview of uncertainty and modelling precision in the context of the District's ability to predict the response of the hydrologic systems which the District strives to manage. A broad range of sensitivity analysis tools are available to explore, display, and quantify the uncertainty in predictions made with hydrologic models of key output variables and system performance indices. In particular, relatively simple deterministic sensitivity analysis methods were discussed, as well as more sophisticated first-order uncertainty analysis and Monte Carlo Sampling methods.

Because of the complexity of the SFWMM and NSM models employed by the District, Monte Carlo analyses may be a very major and unattractive undertaking. Therefore it is recommended that the District begin with the relatively simple deterministic sensitivity analysis procedures, as well as employing the probabilistically based first-order uncertainty analysis method to quantify the uncertainty in key output variables and system performance indices, and the relative contributions of uncertainty in different input variables to the uncertainty in different output variables and system performance indices. These relative contributions will change depending upon which output variables and indices are of interest.

A sensitivity analysis procedure was proposed that would provide a systematic assessment of the impact of parameter errors on the uncertainty in output variable values and performance indices, and of the relative contribution of errors in different parameter values to that output uncertainty. Once the key errors are identified, it should be possible to determine the extend to which parameter uncertainty can be reduced through field investigations, development of better models, and other efforts.

Another component of that program would be the development of automated model calibration procedures. Such procedures can be applied to individual catchments and subsystems, as well as composite systems. These automated calibration procedures have several advantages including the explicit identification of an appropriate statistical objective function, identification of those parameters that best reproduce the calibration data set with the given objective function, and the estimations of the statistical precision of the estimated parameters.

All of these tasks together represent a formidable effort. However, knowledge of the uncertainty associated with model predictions can be as important to management decision and policy formulation as are the predictions themselves. It is clear the District recognizes the importance of knowing more about how good various model predictions are before making decisions based on these predictions.

References

Benjamin J.R., and C.A. Cornell, *Probability, Statistics and Decision for Civil Engineers,* McGraw-Hill Book Co., New York, 1970.

Binley, A.M., and K.J. Beven, Physically-based modelling of catchment hydrology: a likelihood approach to reducing predictive uncertainty, in *Computer Modelling in the Environmental Sciences,* D.G. Farmer and M.J. Rycroft (ed.), Clarendon Press, Oxford, 1991.

Beven, K., Changing Ideas in Hydrology – The Case of Physically-based Models, *J. of Hydrol., 105,* 157-172, 1989.

Beven, K., Prophecy, reality and uncertainty in distributed hydrologic modelling, *Adv. in Water Res., 16,* 41-51, 1993.

Beck, M.B., Water Quality Modeling: A review of the Analysis of Uncertainty, *Water Resour. Res., 23*(8), 1393-1442, 1987.

Covello, V. T., Decision Analysis and Risk Management Decision-Making: Issues and Methods, *Risk Analysis, 7,* 1987, pp. 131-139.

DeGarmo, E.P., W.G. Sullivan, and J.A. Bontadelli, *Engineering Economy,* MacMillian, New York, 1993.

Devore, J., *Probability and Statistics for Engineering and the Sciences,* 3rd Edition, Brooks/Cole Publishing Co., 1991.

Duan, Q., S. Soroosian, and R.P. Ibbitt, A maximum likelihood criterion for use with data collected at unequal time intervals, *Water Resources Research, 24*(7), 1163-1173, 1988.

Eschenback, T.G., Spider plots versus Tornado Diagrams for Sensitivity Analysis, *Interfaces, 22,* 40-46, 1992.

Gan, T.Y., and S.J. Burges, An assessment of a conceptual rainfall-runoff model's ability to represent dynamics of small hypothetical catchments, 2, Hydrologic responses for normal and extreme rainfall, *Water Resources Research 27*(7), 1605-1619, 1990.

Gaven, D.C., and S.J. Burges, Approximate error bounds for simulated Hydrograph, *J. Hydraulic Engineering, 107*(11), 1519-1534, 1981.

Gupta, V.K., and S. Sorooshian, The automatic calibration of conceptual watershed models using derivative-based optimization algorithms, *Water Resources Research, 21*(4), 473-485, 1985a.

Gupta, V.K., and S. Sorooshian, The relationship between data and the precision of parameter estimates of hydrologic models, *Journal of Hydrology 81*, 57-77, 1985b.

Hashimoto, T., J.R. Stedinger and D.P. Loucks, Reliability, Resiliency and Vulnerability Criteria for Water Resource System Performance Evaluation, *Water Resources Res., 18*(1), 14-20, 1982.

Hashimoto, T., D.P. Loucks and J.R. Stedinger, Robustness of Water Resource Systems, *Water Resources Research, 18*(1), 21-26, 1982.

Harlin, J., and C-S. Kung, Parameter Uncertainty in Simulation of Design of Floods in Sweden, J. of Hydrol., 137, 2009-230, 1992.

Hendrickson, J.D., S. Sorooshian and L.E. Brazil, Comparison of Newton-type and direct-search algorithms for calibration of conceptual rainfall-runoff models, *Water Research Research, 24*(5), 691-700, 1988.

Kuczera, G., On the Validity of First-order Prediction Limits for Conceptual Hydrologic Models, *J. of Hydrol., 103,* 229-247, 1988.

Loucks, D.P., J.R. Stedinger and D.A. Haith, *Water Resource Systems Planning and Analysis,* Prentice-Hall, Englewood Cliffs, NJ, 1981.

Schweppe, F.C., *Uncertain Dynamic Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1973.

Sorooshian, S., V.K. Gupta, and J.L. Fulton, Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models--influence of calibration data variability and length on model credibility, *Water Resources Research, 19*(1), 251-259, 1983.

Sorooshian, S., Q. Duan, and V.K. Gupta, Calibration of Rainfall-Runoff Models: Application of global optimization to the Sacramento soil moisture accounting model, *Water Resources Research, 29*(4), 1185-94, 1993.

Stedinger, J.R., and M.R. Taylor, Synthetic Streamflow Generation, Part II. Effect of Parameter Uncertainty, *Water Resources Research, 18*(4), 919-924, 1982.

van Straten, G., Maximum likelihood estimation of parameters and uncertainty in phytoplankton models, in *Uncertainty and Forecasting of Water Quality*, M.B. Beck and G. van Straten (ed.), Springer Verlag, New York, 1983.

Vollenweider, R.A., Advances in Defining Critical Loading Levels for Phosphorus in Lake Eutrophication, *Memorie dell'Istituto Italiano di Idrobiologia, 33*, 53-83, 1976.